(51) International Patent Classification⁷: **C12Q 1/00**

(21) International Application Number: PCT/US01/03791

(22) International Filing Date: 2 February 2001 (02.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/498,098   4 February 2000 (04.02.2000)   US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US   09/498,098 (CIP)
Filed on   4 February 2000 (04.02.2000)

(71) Applicant *(for all designated States except US)*: AURORA BIOSCIENCES CORPORATION [US/US]; 11010 Torreyana Road, San Diego, CA 92121 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: STACK, Jeffrey, H. [US/US]; 6619 Edmonton Avenue, San Diego, CA 92122 (US). WHITNEY, Michael [US/US]; 5171 Frink Avenue, San Diego, CA 92117 (US). CUBITT, Andrew, B. [US/US]; 11083 Caminito Encanto, San Diego, CA 92131 (US). POLLOK, Brian, A. [US/US]; 18283 Sun Maiden Court, San Diego, CA 92127 (US).

(74) Agent: HAILE, Lisa, A.; Gray Cary Ware & Freidenrich LLP, Suite 1600, 4365 Executive Drive, San Diego, CA 92121 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(54) Title: METHODS OF PROTEIN DESTABILIZATION AND USES THEREOF

(57) Abstract: This invention is directed towards methods of destabilizing proteins in living cells, and their use for the development of novel assays. In one embodiment, the invention comprises the use of non-cleavable multimerized ubiquitin fusion proteins to destabilize a target protein, such as a reporter moiety. In one aspect of this method the constructs also comprises a linker that operatively couples the reporter moiety to the multimerized ubiquitin fusion protein. In this embodiment, enzymatic modification of the linker results in a modulation of the coupling of the reporter protein to the multimerized ubiquitin domains resulting in a change in the stability of the reporter moiety. The level of the reporter moiety in the cell can then be used as a measure of the enzymatic activity in the cell. In another embodiment the invention provides for a generalized way of coordinately regulating the cellular concentration of a plurality of target proteins. In one aspect of this method, the target proteins are operatively coupled to a ubiquitin fusion protein via linker containing a protease cleavage site. Cleavage of the linker by a protease results in uncoupling of the target protein from the multimerized ubiquitin construct, and results in an increase in the stability and concentration of the target protein.

METHODS OF PROTEIN DESTABILIZATION AND USES THEREOF

FIELD OF THE INVENTION

The present invention is in the field of protein analysis and more particularly methods of destabilizing proteins and using the destabilized proteins for novel cell based assays.

BACKGROUND OF THE INVENTION

While genomic programs provide ever more sophisticated information on the sequence and patterns of expression of mammalian genes, it is increasingly recognized that integrating this information into a functional model of how a cell works requires an understanding of how the protein products of expressed genes interact within the cell. Although we have made significant improvements in our ability to clone, sequence and analyze DNA sequences, our reciprocal abilities for studying RNA and protein molecules are significantly less facile or advanced. Furthermore, proteins themselves represent significantly more complex molecules in terms of composition, shape and activity compared to double stranded DNA. A central challenge facing workers in the field today is to understand out how a protein's activity and function within a cell are regulated and coordinated within the native physiological context.

Traditionally, genetic analysis has been used for determining the function of gene products and how they interact with other proteins within a common pathway. Unfortunately genetic analysis in vertebrate organisms is extremely time consuming and expensive. An alternative approach is to devise an assay system for a given protein and then screen for compounds that activate or inhibit its function. These compounds can be used to dissect the cellular pathways the protein functions in, as well as serving as potential compounds of therapeutic value.

Although there is tremendous interest in understanding the regulation and interactions of proteins within cells there are relatively few methods that are robust, simple to use, amenable to high throughput screening or can be used effectively within living cells. Furthermore in many cases where specific assays do exist these are restricted in scope to individual enzymatic steps or to one or two defined pathways.

A need thus exists for sensitive methods of interfacing the functional modifications of proteins with optical signals that can be used detect and monitor these changes, for example for use in high throughput screening. In drug screening applications these assays can be applied to find useful compounds that are specific and selective for a particular protein or signal transduction or metabolic pathway.

Proteins may undergo a huge variety of post-translational modifications subsequent to their synthesis in the cell. In many cases these modifications can play critical roles in the functioning and stability of the modified proteins. For example, proteolysis, phosphorylation, covalent attachment of a lipid or lipid derivative, disulfide bond formation, glycosylation and

5   oxidation all can have important functional effects. Many other examples also exist and may play important functional roles within a cell for defined proteins.

One approach to developing a generic assay capable of detecting these myriad post-translational modifications is to operatively couple these activities through a central pathway of protein modification that can be sensitively measured with a common reporter system. In the

10   present invention, the inventors have recognized that by coupling post-translational activities to the stability of a high sensitivity reporter moiety it is possible to develop uniform cell based assays for a range of activities. Importantly these measurements are robust enough for high throughput screening applications, readily adaptable to a range of activities and provide cellular assays that provide information within a living cell.

15   In the present invention, post-translational activities can be measured by providing one or more constructs in which the activity to be measured influences the stability of a reporter moiety. In one embodiment, this may be achieved by providing a reporter moiety that is operatively coupled to a multimerized destabilization domain via a linking moiety. The linking moiety comprises a recognition motif for the target activity such that modification of the linker by the

20   activity results in altered stability of the reporter moiety. If the reporter moiety is an enzymatic reporter gene the method provides a high sensitivity readout that is generally applicable to a range of activities which are otherwise difficult to measure within living cells. The multimerized destabilization domain described herein provides a key advantage in the method because it enables the degree of destabilization to be predictably tuned to any activity level or intrinsic

25   stability of the target protein or reporter moiety.

The regulation of protein stability is an area of particular interest because of its increased recognition as a key regulator of a protein's concentration and function in the cell. Although our knowledge of the factors that control protein stability have grown dramatically in recent years, it is clear that a variety of cellular pathways and environmental cues participate in and control a

30   protein's fate. For example, mis-folding, proteolysis, oxidation and some conformational changes that expose significant surface hydrophobicity readily contribute to the recognition of a protein by the cellular machinery for protein degradation. The majority of cytoplasmic protein degradation involves the ubiquitination of the target protein followed by binding and degradation by the proteasome. (For review see Hershko and Ciechanover (1998) Annu. Rev. Biochem. 67

35   425-79)

A key step in protein ubiquitination, and degradation, is recognition of the target protein by ubiquitin protein ligase or E3 enzyme. This class of enzymes is responsible for the covalent attachment of ubiquitin to the target protein via an amide isopeptide linkage to an -amino group of one of the substrate protein's lysine residues. There are currently believed to be multiple families of E3 enzymes, additionally there is increasing evidence that some E3 proteins exist as multi-subunit protein complexes (Laney and Hochstrasser (1999) Cell 97 427-430). E3 proteins and their associated complexes are believed to be largely responsible for recognizing and ubiquitinating damaged proteins as well as specific destabilization domains present in target proteins. Once recognized, a protein target that has been modified by the addition of a single ubiquitin domain, becomes a substrate for further ubiquitination, either at different sites in the substrate protein, or through extension of the conjugated ubiquitin. This process can thus lead to a poly-ubiquitinated protein with numerous branched ubiquitin domains attached. Once poly-ubiquitinated, the protein is recognized with high affinity by the proteasome where it is degraded.

The addition of specific destabilization domains to a target protein has in some cases been demonstrated to destabilize that target protein. A key challenge in this area has been to provide a predictable way of creating graded levels of destabilization for a given protein that that can be utilized in manipulating the steady state levels or dynamic temporal regulation of that protein. The present inventors have discovered for the first time that by providing stable multimerized linear chains of individual destabilization domains, such as ubiquitin, it is possible to create a generic method of protein destabilization that is widely applicable to virtually any protein. Importantly, this approach has the advantage that the degree of destabilization can be accurately controlled by varying the number of destabilization domains added to the target protein. As a result, the actual cellular concentration and half-life of an exogenously expressed protein in a cell or living organism can be accurately and reproducibly controlled. By coupling 1, 2, 3, 4 or more copies of ubiquitin to the reporter gene -lactamase it has been possible to regulate the protein concentration of this protein in the cell over a 10-fold range compared to the native protein. The present inventors have applied this discovery to create an assay technology that is broadly capable of measuring a wide range of post-translational activities.

SUMMARY OF THE INVENTION

This invention provides a fluorescent, bioluminescent or enzymatic substrate useful as an optical probe or sensor of post-translational modifications, such as proteolysis. In one embodiment, the invention comprises a reporter moiety that is functionally coupled to one or

more destabilizing domains via a linker. The linker typically contains a recognition motif for an activity. Modification of the linker by the activity results in uncoupling of the reporter moiety from the destabilizing domain(s) with a corresponding change in the stability of the reporter moiety. The level of activity within a sample is sensed by a measurable change in the level of the reporter moiety, for example by detecting at least one optical property of the reporter moiety, or by detecting at least one optical property of detectable product of the reporter moiety. FIG. 1.

In one embodiment the reporter moiety is an enzymatic reporter such as alkaline phosphatase, -galactosidase, chloramphenicol acetyltransferase, -glucuronidase, peroxidase, -lactamase, bioluminescent proteins, luciferases and catalytic antibodies. In another embodiment the reporter moiety is a naturally fluorescent protein, epitope or structural protein.

In one aspect the linker moiety is an amino acid sequence that covalently couples the reporter moiety to the multimerized destabilization domain. In another aspect, the linker moiety comprises two separate amino acid sequences, one of which is covalently coupled to the reporter moiety, and one of which is coupled to the multimerized destabilization domain. Coupling of the reporter moiety to the destabilization domains occurs through the non-covalent interaction or binding of the two amino acid sequences of the linker together. In either case, modification of the linker by the activity results in a modulation of the coupling of the reporter moiety to the multimerized destabilization domains. In one aspect of this method the activity is selected from the group consisting of a protease activity, a protein kinase activity and a phosphoprotein phosphatase activity.

In one aspect the multimerized destabilization domain comprises two, three, four, or more copies of the destabilization domain covalently coupled together in a linear chain. In one embodiment, the destabilization domains comprise ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In another aspect the invention comprises a method of regulating the concentration of one or more target proteins in a cell. The method involves the creation of a fusion protein containing the protein of interest coupled to one or more destabilization domains. In different embodiments the protein of interest may be coupled to a multimerized destabilization domain comprising two or more copies of the destabilization domain. In one embodiment, the destabilization domains comprise ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by -NH-ubiquitin protein

endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In one aspect of this method, the fusion protein may additionally comprise a linker that couples the protein of interest to one or more destabilization domains. The linker typically

5   comprises a protease cleavage site for a protease. Cleavage of the linker by the protease modulates the coupling of the multimerized destabilization domain to the protein of interest, thereby providing a method of rapidly modulating the stability of one or more proteins of interest in the cell simultaneously. The protease may be introduced into the cell, or its activity regulated by the presence of a membrane permeant small molecule inhibitor. In one embodiment of this

10   method, the protease does not naturally occur in the target cell.

In another aspect the invention includes a recombinant DNA molecule, comprising a nucleic acid sequence encoding for one or more destabilization domains, a target protein, and a linker moiety that operatively couples the destabilization domain(s) to the target protein. In different embodiments the protein of interest may be coupled to one, two, three, four or more

15   copies of the destabilization domain. In one embodiment, the destabilization domains comprise ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger

20   or more bulky amino acid.

In another embodiment the invention includes a recombinant protein molecule, comprising an amino acid sequence encoding for one or more destabilization domains, a target protein, and a linker moiety that operatively couples the multimerized destabilization domain to the target protein.

25   In another aspect the invention includes a cell or transgenic organism comprising a nucleic acid sequence encoding for a one or more destabilization domains, a target protein, and a linker moiety that operatively couples the destabilization domain(s) to the target protein. In different embodiments the protein of interest may be coupled to one, two, three, four or more copies of the destabilization domain. In one embodiment, the destabilization domains comprise

30   ubiquitin, or a homolog thereof. In a preferred embodiment the multimerized copies of ubiquitin are not cleavable by -NH-ubiquitin protein endoproteases. In one embodiment the ubiquitin domains comprise a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases. In one aspect of this embodiment the mutation represents the mutation of glycine 76 to a larger or more bulky amino acid.

In another embodiment the invention includes a method for identifying a modulator of an activity, comprising the use of the inventions cells or transgenic organisms. The method includes contacting the cells with a test chemical and detecting the activity of the reporter moiety. Additional claims involve the steps of contacting the cell with an activator of the activity prior to

5   the addition said test chemical, and of in parallel determining the cell viability of the cell in the presence of the test chemical.

In another embodiment the invention is directed to the test chemical and a pharmaceutical composition comprising a test chemical identified by the methods of the present invention.

The accompanying drawings, which are incorporated in and form part of the

10   specification, merely illustrate embodiments of the present invention. Together with the remainder of the specification, they are meant to serve to explain certain principles of the invention to those of skill in the art.

BRIEF DESCRIPTION OF THE FIGURES

15

FIG. 1 General schematic overview of parent construct pcDNA3-UbiquitinG76V-Bla. Shown are important coding regions including the ubiquitin- -lactamase fusion coding region, various promoters and important restriction sites used in the cloning of derivative constructs.

20   **FIG. 2** TNT *in vitro* synthesis and degradation experiments with Met, 1, 2, 3 or 4 copies of ubiquitinG76V fused to -lactamase. The kinetics of turnover *in vitro* in (A) were determined by chase reactions at 37°C and products analyzed by SDS-PAGE. The effect of the proteasome inhibitor MG132 at 50 M in the TNT synthesis reaction is shown in (B).

25   **FIG. 3** Turnover in vitro of labeled fusion proteins of uncleavable ubiquitinG76V fused to GFP. TNT synthesis reactions were incubated in chase lysate at 37°C and products analyzed by SDS-PAGE.

FIG. 4 Turnover reactions *in vitro* of labeled uncleavable ubiquitin caspase-3 fusions. TNT

30   reactions were incubated in chase lysate at 37°C and products analyzed by SDS-PAGE.

FIG. 5 FACS™ analysis of uncleavable ubiquitin -lactamase fusions. Jurkat cells expressing ubiquitinG76V-Bla fusion proteins were analyzed for -lactamase expression by flow cytometry. The R5+R6+R7 region was designated as Bla[+] and the percentage of cells in that

35   region is shown in the bar graph.

FIG. 6 Kinetics of degradation *in vivo* of ubiquitinG76V- -lactamase fusion proteins. Jurkat cells expressing the various ubiquitinG76V-Bla fusions were treated with cycloheximide to initiate a chase and aliquots of cells were removed at the indicated times. The cells were lysed and the -lactamase activity in the lysates was determined by an *in vitro* reaction using the fluorescent substrate CCF2. The -lactamase activity was measured by cleavage of CCF2 and represented as emission at 460nm.

FIG. 7 Caspase cleavage of 2XUb-DEVD-Bla results in the stabilization of -lactamase. TNT synthesis reactions were performed to generate labeled fusion proteins of the caspase substrate 2XUb-DEVD-Bla and control 2XUb-DEVA-Bla. In (A), the labeled proteins were incubated with purified caspase-3 to verify that 2XUb-DEVD-Bla can be cleaved by caspase-3 and 2XUb-DEVA-Bla cannot. In (B), the products of the caspase-3 cleavage reactions were incubated with chase extract and samples analyzed by SDS-PAGE.

FIG. 8 Dose-response curves for an inducer and an inhibitor of caspase activation with Jurkat cells expressing 2XUb-DEVD-Bla. Varying concentrations of antiFas IgM were incubated with 2XUb-DEVD-Bla-expressing Jurkat cells for 6 hours at 37°C and caspase activity was measured following a cycloheximide chase to clear uncleaved reporters. The cells were loaded with CCF2-AM and -lactamase activity measured and expressed as a 460/530nm ratio. Jurkat cells expressing 2XUb-DEVD-Bla were treated with varying concentrations of the caspase inhibitor ZVAD-fmk and then treated with 75 ng/ml antiFas IgM. The cells were incubated for 6 hours at 37oC, cycloheximide for 1 hour at 37°C and -lactamase activity measured using CCF2-AM as described above.

FIG. 9 *In vitro* cis-cleavage activity of UbiquitinG76V-HRV 2A-Bla fusions. Labeled UbiquitinG76V-HRV 2A protease -lactamase fusions were produced in TNT reactions and then analyzed by SDS-PAGE. (A) shows that the cis-cleavage of HRV-Bla fusions is blocked by mutation of putative catalytic residues (C106 and D35). (B) The TNT reactions were incubated in chase extract to show the selective stabilization of the cleavage product.

FIG. 10 Rapid degradation of 2XUb-Bla *in vitro* requires polyubiquitination and proteasome activity. TNT synthesis reactions were incubated in chase extract containing the indicated inhibitors for 20 minutes at 37°C. MG132 and ALLN were present at 50 M, lactacystin at 10 mM and MeUb at 200 g/ml.

7

FIG. 11 Dose-response curves for proteasome inhibitors on Jurkat cells expressing 2XUb-Bla reporter. Cells were treated with varying concentrations of MG132 or ALLN for 30 minutes and then cycloheximide was added and the cells incubated at 37°C for one hour. The cells were loaded with CCF2-AM to measure -lactamase activity as described above.

DETAILED DESCRIPTION

Definitions

The techniques and procedures are generally performed according to conventional methods in the art and various general references. (Lakowicz, J.R. Topics in Fluorescence Spectroscopy, (3 volumes) New York: Plenum Press (1991), and Lakowicz, J. R. (1996) Scanning Microsc Suppl. 10 213-24, for fluorescence techniques; Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., for molecular biology methods; Cells: A Laboratory Manual, 1st edition (1998) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., for cell biology methods; Optics Guide 5 Melles Griot® Irvine CA, and Optical Waveguide Theory, Snyder & Love published by Chapman & Hall for general optical methods, which are incorporated herein by reference which are provided throughout this document).

"Activity" refers to the enzymatic or non-enzymatic activity capable of modifying an amino acid residue or peptide bond (preferably enzymatic). Such covalent modifications include proteolysis, phosphorylation, dephosphorylation, glycosylation, methylation, sulfation, prenylation and ADP-ribsoylation. The term includes non-covalent modifications including protein-protein interactions, and the binding of allosteric, or other modulators or second messengers such as calcium, or cAMP or inositol phosphates to a polypeptide.

Amino acid "substitutions" are defined as one for one amino acid replacements. They are conservative in nature when the substituted amino acid has similar structural and/or chemical properties. Examples of conservative replacements are substitution of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine.

Amino acid "insertions" or "deletions" are changes to or within an amino acid sequence. They typically fall in the range of about 1 to 5 amino acids. The variation allowed in a particular amino acid sequence may be experimentally determined by producing the peptide synthetically or by systematically making insertions, deletions, or substitutions of nucleotides in the gene sequence using recombinant DNA techniques.

"Animal" as used herein may be defined to include human, domestic (cats, dogs, etc), agricultural (cows, horses, sheep, goats, chicken, fish, etc) or test species (frogs, mice, rats, rabbits, simians, etc).

"Chimeric" molecules are polynucleotides or polypeptides which are created by combining one or more of nucleotide sequences of this invention (or their parts) with additional nucleic acid sequence(s). Such combined sequences may be introduced into an appropriate vector and expressed to give rise to a chimeric polypeptide which may be expected to be different from the native molecule in one or more of the following characteristics: cellular location, distribution, ligand- binding affinities, interchain affinities, degradation/turnover rate, signaling, etc.

The terms "cleavage site" or "protease site" refers to the bond cleaved by the protease (e.g. a scissile bond) and typically the surrounding three to four amino acids of either side of the bond.

"Control elements" or "regulatory sequences" are those non-translated regions of the gene or DNA such as enhancers, promoters, introns and 3' untranslated regions which interact with cellular proteins to carry out replication, transcription, and translation. They may occur as boundary sequences or even split the gene. They function at the molecular level and along with regulatory genes are very important in development, growth, differentiation and aging processes.

"Corresponds to" refers to a polynucleotide sequence that is' homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to all or a portion of a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

"Derivative" refers to those polypeptides which have been chemically modified by such techniques as ubiquitination, labeling, pegylation (derivatization with polyethylene glycol), and chemical insertion or substitution of amino acids such as ornithine which do not normally occur in human proteins.

A "destabilization domain" refers to a protein, polypeptide or amino acid sequence that is capable of modulating the stability of a protein of interest when functionally coupled to the protein of interest. Examples of destabilizing domains include ubiquitin, PEST sequences, cyclin destruction boxes and hydrophobic stretches of amino acids. Preferred destabilization domains include ubiquitin and homologs thereof, particularly those comprising mutations that prevent, or significantly reduce, the cleavage of ubiquitin multimers by -NH-ubiquitin protein

endoproteases. Examples of such mutations include the mutation of glycine 76 to another amino acid, particularly an amino acid selected from the group consisting of Ala, Leu, Ile, Phe, Tyr, Val, Met, Cys, His, Trp, Pro, Arg, Lys, Thr and Ser. Preferred is UbiquitinG76V.

A "detectable product" is a chemical moiety used for detecting a reporter moiety. They include, but are not limited to, radionuclides, enzymes, fluorescent, chemi-luminescent, or chromogenic agents. Detectable products associate with, establish the presence of, and may allow quantification of a particular nucleic sequence, amino acid sequence or reporter moiety. Preferred detectable products are retained within living cells and provide a fluorescence readout that is compatible with fluorescent activated cell sorting (FACS) analysis.

The term "engineered protease site" refers to a protease site that has been modified from the naturally existing sequence by at least one amino acid substitution.

The term "homolog" refers to two sequences or parts thereof, that are greater than, or equal to 85% identical when optimally aligned using the ALIGN program. Homology or sequence identity refers to the following. Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred. Alternatively and preferably, two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M.O., (1972) in Atlas of Protein Sequence and Structure 5, National Biomedical Research Foundation, 101-110, and Supplement 2 to this volume, pp. 1-10.

An "inhibitor" is a substance that retards or prevents a chemical or physiological reaction or response. Common inhibitors include but are not limited to antisense molecules, antibodies, antagonists and their derivatives.

"Isolated" refers to material removed from its original environment (e.g. the natural environment if it is naturally occurring), and thus is altered from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

The term "linker" or "linker moiety" refers to an amino acid, polypeptide or protein sequence that serves to operatively couple a reporter moiety to one or more destabilization domains. Linkers may comprise a single polypeptide chain that covalently couples the reporter

moiety to the multimerized destabilization domain. Alternatively the linker may comprise two separate polypeptides. Typically the first polypeptide is covalently coupled to the reporter moiety, and the second polypeptide is covalently coupled to the multimerized destabilization domain. Generally the first and second polypeptides comprising the linker moiety in this

5   embodiment are capable of interacting or associating such that the interaction or association operatively couples the reporter moiety to the multimerized destabilization domain. Preferably the linker moiety is non-cleavable by -NH-ubiquitin protein endoproteases. Linkers may be of any size.

The term "modulates" refers to, either the partial or complete, enhancement or inhibition

10   (e.g. attenuation of the rate or efficiency) of an activity or process.

The term "modulator" refers to a chemical compound (naturally occurring or non-naturally occurring), such as a biological macromolecule (e.g., nucleic acid, protein, non-peptide, or organic molecule), or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian, including human) cells or tissues. Modulators are evaluated

15   for potential activity as inhibitors or activators (directly or indirectly) of a biological process or processes (e.g., agonist, partial antagonist, partial agonist, inverse agonist, antagonist, antineoplastic agents, cytotoxic agents, inhibitors of neoplastic transformation or cell proliferation, cell proliferation-promoting agents, and the like) by inclusion in screening assays described herein. The activity of a modulator may be known, unknown or partially known.

20   The term "multimerized destabilization domain" refers to at least two destabilization domains that are linearly coupled together. Preferred multimerized domains are non-cleavable by -NH-ubiquitin protein endoproteases. The term does not include naturally occurring poly-ubiquitin chains in which the ubiquitin monomers are coupled together via isopeptide bonds attached to the -amino group of lysine. The term also does not include naturally occurring

25   multi-ubiquitin genes, are cleavable by -NH-ubiquitin protein endoproteases to create ubiquitin monomers. The destabilization domains present in the multimerized destabilization domain are typically the same, but need not necessarily be identical.

"Naturally fluorescent protein" refers to proteins capable of forming a highly fluorescent, intrinsic chromophore either through the cyclization and oxidation of internal amino acids within

30   the protein or via the enzymatic addition of a fluorescent co-factor. Typically such chromophores can be spectrally resolved from weakly fluorescent amino acids such as tryptophan and tyrosine.

"Naturally occurring" refers to a polypeptide produced by cells which have not been genetically engineered or which have been genetically engineered to produce the same sequence

35   as that naturally produced. Specifically contemplated are various polypeptides that arise from

post-transnational modifications. Such modifications of the polypeptide include but are not limited to acetylation, carboxylation, glycosylation, phosphorylation, lipidation, proteolytic cleavage and acylation.

An "oligonucleotide" or "oligomer" is a stretch of nucleotide residues which has a sufficient number of bases to be used in a polymerase chain reaction (PCR), a site directed mutagenesis reaction or a cassette to create a desired sequence element. These short sequences are based on (or designed from) genomic or cDNA sequences and are used to amplify, mutate or create particular sequence elements. Oligonucleotides or oligomers comprise portions of a DNA sequence having at least about 10 nucleotides and as many as about 50 nucleotides, preferably about 15 to 30 nucleotides. They are chemically synthesized and may also be used as probes.

An "oligopeptide" is a short stretch of amino acid residues and may be expressed from an oligonucleotide. It may be functionally equivalent to and either the same length as or considerably shorter than a "fragment ", "portion ", or "segment" of a polypeptide. Such sequences comprise a stretch of amino acid residues of at least about 5 amino acids and often about 17 or more amino acids, typically at least about 9 to 13 amino acids, and of sufficient length to display biologic and/or immunogenic activity.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The term "operably coupled" refers to a juxtaposition wherein the components so described are either directly or indirectly coupled. Examples of directly coupled components include proteins that are translationally fused together. Examples of indirectly coupled components include proteins that can functionally associate either transiently, or persistently, through a binding interaction.

The term "polynucleotide" refers to a polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides. Modified forms and analogs of either type of nucleotide are also included, as are ribonucleotides or deoxynucleotides linked via novel bonds such as those described in U.S. Patent No. 5,532,130, European Patent Applications EP 0 839 830, EP 0 742 287, EP 0 285 057 and EP 0 694 559. The term includes single and double stranded forms of nucleotides, or a mixture of single and double stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for

example, tritylated bases and unusual bases such as inosine, as well as other chemical or enzymatic modifications.

The term "polypeptide" refers to a amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modification include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristolyation, oxidation, pergylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to protein such as arginylation. (See Proteins- Structure and Molecular Properties 2nd Ed., T.E. Creighton, W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Pres, New York, pp. 1-12 (1983).

A "portion" or "fragment" of a polynucleotide or nucleic acid comprises all or any part of the nucleotide sequence having fewer nucleotides than about 6 kb, preferably fewer than about 1 kb which can be used as a probe. Such probes may be labeled with reporter molecules using nick translation, Klenow fill-in reaction, PCR or other methods well known in the art. After pretesting to optimize reaction conditions and to eliminate false positives, nucleic acid probes may be used in Southern, northern or in situ hybridizations to determine whether DNA or RNA encoding the protein is present in a biological sample, cell type, tissue, organ or organism.

"Probes" are nucleic acid sequences of variable length, preferably between at least about 10 and as many as about 6,000 nucleotides, depending on use. They are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. They may be single- or double-stranded and carefully designed to have specificity in PCR, hybridization membrane-based, or ELISA-like technologies.

13

The term "recognition motif" refers to all or part of a polypeptide sequence recognized by a post-translational modification activity to enable a polypeptide to become modified by that post-translational modification activity. Typically, the affinity of a protein, e.g. enzyme, for the recognition motif is about 1 mM (apparent K_d), preferably a greater affinity of about 10 M,

5   more preferably, 1 M or most preferably has an apparent K_d of about 0.1 M. The term is not meant to be limited to optimal or preferred recognition motifs, but encompasses all sequences that can specifically confer substrate recognition to a peptide. In some embodiments the recognition motif is a phosphorylated recognition motif (e.g. includes a phosphate group), or comprises other post-translationally modified residues.

10   "Recombinant nucleotide variants" are polynucleotides that encode a protein. They may be synthesized by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively.

15   "Recombinant polypeptide variant" refers to any polypeptide which differs from a naturally occurring polypeptide by amino acid insertions, deletions and/or substitutions, created using recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing characteristics of interest may be found by comparing the sequence of a polypeptide with that of related polypeptides and minimizing the

20   number of amino acid sequence changes made in highly conserved regions.

A "reporter moiety" includes any protein that directly or indirectly produces a specific detectable product, or cellular phenotype, such as drug resistance that can be used to monitor transcription of a gene. Preferred reporter moieties include proteins with an enzymatic activity that provides enzymatic amplification of gene expression such as alkaline phosphatase, -

25   galactosidase, chloramphenicol acetyltransferase, -glucuronidase, peroxidase, -lactamase, bioluminescent proteins, luciferases and catalytic antibodies. Other reporter moieties include proteins such as naturally fluorescent proteins or homologs thereof, cell surface proteins or the native or modified forms of an endogenous protein to which a specific assay exists or can be developed in the future. Preferred reporter moieties for use in the present invention provide for a

30   fluorescent readout that is compatible with fluorescent activated cell sorting (FACS) analysis.

A "signal or leader sequence" is a short amino acid sequence which is or can be used, when desired, to direct the polypeptide through a membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous sources by recombinant DNA techniques.

A "standard" is a quantitative or qualitative measurement for comparison. Preferably, it is based on a statistically appropriate number of samples and is created to use as a basis of comparison when performing diagnostic assays, running clinical trials, or following patient treatment profiles. The samples of a particular standard may be normal or similarly abnormal.

5   The term "stringent hybridization conditions", refers to an overnight incubation at 42 °C in a solution comprising 50 % formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10 % dextran sulfate and 20 g/ml denatured sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65 °C. Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the
10   present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lower stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37 °C in a solution comprising 6x SSPE (20X SSPE=3M NaCl; 0.2M NaH2PO4;
15   0.02M EDTA, pH 7.4), 0.5% SDS, 30 % formamide, 100 g/ml salmon sperm blocking DNA; followed by washes at 50 °C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC). Variation in the above conditions may be accomplished through the inclusion and / or substitution of alternative blocking reagents used to suppress background
20   in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility. A polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in
25   the sequence listing), or to a complementary stretch of T (or U) residues would not be included in the definition of a "polynucleotide" since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch, or the complement thereof.

The term "target" refers to a biochemical entity involved a biological process. Targets are typically proteins that play a useful role in the physiology or biology of an organism. A
30   therapeutic chemical binds to target to alter or modulate its function. As used herein, targets can include cell surface receptors, G-proteins, kinases, ion channels, phopholipases, proteases and other proteins mentioned herein.

The term "test chemical" refers to a chemical to be tested by one or more screening method(s) of the invention as a putative modulator. A test chemical can be any chemical, such
35   as an inorganic chemical, an organic chemical, a protein, a peptide, a carbohydrate, a lipid, or a

15

combination thereof. Usually, various predetermined concentrations of test chemicals are used for screening, such as 0.01 micromolar, 1 micromolar and 10 micromolar. Test chemical controls can include the measurement of a signal in the absence of the test compound or comparison to a compound known to modulate the target.

5   The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage identical to a sequence", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or may comprise a complete

10   cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two

15   (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the

20   polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm

25   of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of

30   homology over the comparison window) generated by the various methods selected. The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage identical to a sequence" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g.,

35   A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing

the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 30 percent sequence identity, preferably at least 50 to 60 percent sequence identity, more usually at least 60 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 30 percent sequence identity, preferably at least 40 percent sequence identity, more preferably at least 50 percent sequence identity, and most preferably at least 60 percent sequence identity. Preferably, residue positions which are not identical differ by conservative amino acid substitutions. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamic-aspartic, and asparagine-glutamine.

Since the list of technical and scientific terms cannot be all encompassing, any undefined terms shall be construed to have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. Furthermore, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, reference to a "restriction enzyme" or a "high fidelity enzyme" may include mixtures of such enzymes and any other enzymes fitting the stated criteria, or reference to the method includes reference to one or more methods for obtaining cDNA sequences which will be known to those skilled in the art or will become known to them upon reading this specification.

Before the present sequences, variants, formulations and methods for making and using the invention are described, it is to be understood that the invention is not to be limited only to

the particular sequences, variants, formulations or methods described. The sequences, variants, formulations and methodologies may vary, and the terminology used herein is for the purpose of describing particular embodiments. The terminology and definitions are not intended to be limiting since the scope of protection will ultimately depend upon the claims.

5

I.   MULTIMERIZED DESTABILIZATION DOMAINS

10   Destabilization domains include proteins, protein domains and amino acid sequences that when functionally coupled to a target protein effect a change in the half-life of that protein when expressed in a cell. Examples include PEST domains, stretches of hydrophobic amino acids, phosphorylation dependent degradation signals, cyclin destruction boxes and the addition of ubiquitin domains. Preferred as a destabilization domain is ubiquitin and homologs thereof,
15   particularly mutants or homologs comprising mutations that prevent, or significantly reduce, the cleavage of ubiquitin multimers by -NH-ubiquitin protein endoproteases.   In general, destabilization domains function by causing the target protein to be recognized by one or more elements of the cellular protein degradation apparatus. Once marked for destruction, the protein is actively recruited into the 28S proteasome where the protein is degraded. Within the cell a
20   variety of signals may target a protein for degradation. In some cases a destabilization feature may be revealed in a protein as a result of oxidation, mis-folding or proteolysis. For example, stretches of hydrophobic amino acids are often exposed in denatured or improperly folded proteins thereby targeting them for degradation. Short stretches of hydrophobic amino acids, or hydrophobic domains, also occur in correctly folded proteins and have been identified in proteins
25   with short half lives.
        For example, the Deg 1 domain of yeast mating type transcription factor 2 is a 19 residue element that forms an amphipathic helix with an exposed hydrophobic face, and is responsible for the rapid degradation of this protein (Johnson et al., (1998) Cell 94 217-227). These elements are believed to be recognized by E3 ubiquitin ligases and target the protein to
30   degradation through the ubiquitin system described below.
        PEST domains (regions rich in the amino acids proline (P), glutamic acid (E), serine (S) and threonine (T)) are often located at the C-terminal domains of relatively unstable proteins. (Rogers, et al., (1986) Science 234 (4774) 364-8). A well characterized PEST domain is located in residues 422 to 461 of ornithine decarboxylase, and has been used to successfully destabilize a
35   number of proteins including the green fluorescent protein from *Aequorea* green fluorescent

protein (*Li et al.* J. Biol. Chem. (1998) 273 (52) 34970-5). Certain PEST sequences are believed to be recognized by the 26S proteasome subunit directly and do not require ubiquitination.

PEST sequences may also be regulated by phosphorylation, for example multiple phosphorylation within the PEST sequences of the yeast G1 cyclins Cln3 and Cln2 are required
5 for degradation.

Phosphorylation dependent degradation signals have also been identified in the transcription factors NF- B and -catenin, in addition to many cell cycle regulatory proteins such as cyclins. (Ghosh *et al.*, (1998) Ann. Rev. Immunol. 16 225-260; Aberle *et al.*, (1997) EMBO J. 16 3797-3804; Koepp *et al.*, (1999) 97 431-434). These proteins include
10 phosphorylation dependent recognition sequences that bind to one of the growing family of E3 ubiquitin ligases only when the site is phosphorylated. In NF- B, the binding domain for the E3 ubiquitin ligase comprises the relatively short sequence DS*GLDS*, (SEQ. ID. NO.: 1) where S* denotes phosphoserine. Binding to the E3 ubiquitin ligase does not require a ubiquitination conjugation site in this case.

15 The cell-cycle destruction box is a partially conserved 9 amino acid sequence motif usually located approximately 40-50 amino acid residues from the N-terminus of the protein first described for the A and B type cyclins. The consensus destruction box sequence has the general structure as shown in **Table 1** below.

TABLE 1								
Consensus destruction box sequence								
R	(A/T)	(A)	L	(G)	x	(I/V)	(G/T)	(N)
1	2	3	4	5	6	7	8	9

20 Amino acid residues, or combinations of two residues, that appear in parentheses in the above structure occur in more than 50 % of known destruction sequences. The residues at positions 1 and 4 are conserved in all destruction boxes.

Ubiquitin (SEQ. ID. NO.: 2), a 76 amino acid polypeptide found in all eukaryotic cells, is centrally involved in the mechanism of targeting a protein for degradation by the cell. In general,
25 the covalent attachment of a ubiquitin domain (SEQ. ID. NO.: 2), to a protein represents a primary recognition motif for binding of that protein to the proteasome. The attachment of ubiquitin (SEQ. ID. NO.: 2) to the protein typically occurs after recognition of one or more of the destabilization domains discussed above, or some other destabilizing feature of a protein. Attachment of ubiquitin (SEQ. ID. NO.: 2) occurs via the reversible isopeptide linkage of the
30 carboxy-terminus of ubiquitin (SEQ. ID. NO.: 2) to lysine residues in the target protein. After the addition of the first ubiquitin domain (SEQ. ID. NO.: 2), further ubiquitin moieties (SEQ. ID. NO.: 2) may subsequently be added via free lysine residues in ubiquitin (SEQ. ID. NO.: 2) to

create branched poly-ubiquitin chains on the substrate protein. These reactions are catalyzed by a family of enzymes that are often referred to as the ubiquitination complex. Once the target protein comprises one or more copies of ubiquitin (SEQ. ID. NO.: 2) it binds with high affinity to the proteasome where it is degraded. (See generally, Hershko *et al.*, (1998) Annu. Rev. Biochem. 76 425-79; Laney *et al.*, (1999) Cell 97 427-430).

The ubiquitin gene typically comprises multiple copies of the ubiquitin coding sequence (SEQ. ID. NO.: 2). Individual ubiquitin domains (SEQ. ID. NO.: 2) are post-translationally formed from the poly-ubiquitin gene by cleavage of the expressed protein by specific -NH-ubiquitin protein endoproteases that are present within all eukaryotic cells. (Jonnalagadda *et al.*, (1989) J. Biol. Chem. 264 10637-10642. The endoproteases will cleave either multiple ubiquitin – ubiquitin chains, or ubiquitin - fusion protein constructs, provided that the last amino acid of the ubiquitin moiety (SEQ. ID. NO.: 2) is glycine. If this last amino acid is mutated to a more bulky amino acid the ubiquitin fusion protein is not cleavable by -NH-ubiquitin protein endoproteases.

The present inventors have recognized for the first time that the creation of multiple ubiquitin fusion proteins that are not cleavable by the -NH-ubiquitin protein endoproteases provides for a facile and tunable method of regulating protein stability. This invention has many important applications for developing novel assays for intracellular activities, and as a regulatable method of coordinately controlling protein concentrations within the cell.

II.   REPORTER MOIETIES

Enzymatic reporter moieties include any protein capable of catalyzing the creation of a detectable product. Specific examples include alkaline phosphatase, -galactosidase, chloramphenicol acetyltransferase, -glucuronidase, peroxidase, -lactamase, catalytic antibodies, luciferases and other bioluminescent proteins.

Alkaline phosphatase, including human placental and calf intestinal alkaline phosphatase (for example, GenBank Accession # U89937), can be measured using colorimetric, fluorescent and chemiluminescent substrates. (Berger, J., *et al.* (1988) Gene 66 1-10; Kain, S.R. (1997) Methods. Mol. Biol. 63 49-60) Alkaline phosphatase is widely used in transcriptional assays, typically by measuring secreted alkaline phosphatase (SEAP).

-galactosidase (-Gal) the gene product of the bacterial gene LacZ, is also widely used as a reporter gene for transcriptional analysis and may be assayed via histochemical, fluorescent or chemiluminescent substrates, either within intact, or permeabilized cells. (See, U.S. Patent No.

5,070,012, issued Dec, 3, 1991 to Nolan *et al.*, and Bronstein, I., *et al.*, (1989) J. Chemilum. Biolum. 4 99-111).

-glucuronidase (GUS) is widely used for transcriptional analysis in higher plants and may also be assayed using a variety of histochemical and fluorescent substrates. (See generally U.S. Patent No. 5,599,670, issued Feb, 4, 1997 to Jefferson).

Chloramphenicol acetyltransferase (CAT), encoded by the bacterial Tn9 gene, is widely used for transcriptional assays and is traditionally measured using a radioisotopic assay in cell extracts (See Gorman *et al.*, (1982) 2 1044-51).

Catalytic antibodies are also amenable for use as reporter genes, if the reaction catalyzed by the antibody results in the formation of a detectable product. Examples include the aldolase specific antibodies 38C2 and 33F12 that catalyze the synthesis of novel fluorogenic retro-aldol reactions (List *et al.*, (1998) Proc. Natl. Acad. Sci USA 95 15351-15355). Typical antibody substrates are cell permeant nonpolar organic molecules that are not substrates for the natural enzymes and are thus good markers of enzyme activity.

□- *Lactamases*

A large number of -lactamases have been isolated and characterized, all of which would be suitable for use in accordance with the present method. Initially, -lactamases were divided into different classes (I through V) on the basis of their substrate and inhibitor profiles and their molecular weight (Richmond, M. H. and Sykes, R. B., (1973) Adv. Microb. Physiol. 9 31-88). More recently, a classification system based on amino acid and nucleotide sequence has been introduced (Ambler, R.P., (1980) Phil. Trans. R. Soc. Lond. [Ser.B.] 289 321-331). Class A -lactamases possess a serine in the active site and have an approximate weight of 29kd. This class contains the plasmid-mediated TEM -lactamases such as the RTEM enzyme of pBR322. Class B -lactamases have an active-site zinc bound to a cysteine residue. Class C enzymes have an active site serine and a molecular weight of approximately 39kd, but have no amino acid homology to the class A enzymes.

The coding regions of an exemplary ß-lactamase employed in the methods described herein include SEQ. ID. NOs: 3 through 7. Nucleic acids encoding proteins with -lactamase activity can be obtained by methods known in the art, for example, by polymerase chain reaction of cDNA using primers based on a DNA sequence in SEQ. ID. NO.: 3. PCR methods are described in, for example, U.S. Patent No. 4,683,195; Mullis *et al.* (1987) Cold Spring Harbor Symp. Quant. Biol. 51 263; and Erlich, ed., PCR Technology, (Stockton Press, NY, 1989).

Preferably, beta-lactamase polynucleotides encode an intracellular form of a protein with beta-lactamase activity that lacks a functional signal sequence. This provides the advantage of trapping the normally secreted beta-lactamase protein within the cell, which enhances the signal

to noise ratio of the signal associated with beta-lactamase activity, and enables the individual cells to be FACS™ sorted. For example, in any of the polypeptides of SEQ. ID. NO.: 3-7, the signal sequence has been replaced with the amino acids Met-Ser. Accordingly, upon expression, beta-lactamase activity remains within the cell. For expression in mammalian cells it is preferable to use beta-lactamase polynucleotides with nucleotide sequences preferred by mammalian cells. In some applications secreted forms of beta-lactamase can be used with the methods described herein.

A variety of colorimetric and fluorescent substrates of -lactamase are available. Fluorescent substrates include those capable of changes, either individually or in combination, of total fluorescence, excitation or emission spectra or fluorescence resonance energy transfer (FRET), for example those described in U.S. Patent No. 5,741,657, issued April 21, 1998, and U.S. Patent 5,955,604, issued September 22, 1999. Any membrane permanent β-lactamase substrate capable of being measured inside the cell after cleavage can be used in the methods and compositions of the invention. Membrane permanent β-lactamase substrates will not require permeablizing eukaryotic cells either by hypotonic shock or by electroporation. Generally, such non-specific pore forming methods are not desirable to use in eukaryotic cells because such methods injure the cells, thereby decreasing viability and introducing additional variables into the screening assay (such as loss of ionic and biological contents of the shocked or electroplated cells). Such methods can be used in cells with cell walls or membranes that significantly prevent or retard the diffusion of such substrates. Preferably, the membrane permeant β-lactamase substrates are transformed in the cell into a β-lactamase substrate of reduced membrane permeability (usually at least five-fold less permeable) or that is membrane impermeant. Transformation inside the cell can occur via intracellular enzymes (e.g. esterases) or intracellular metabolites or organic molecules (e.g. sulfhydryl groups).

Bioluminescent proteins

Preferred bioluminescent proteins include firefly, bacterial or click beetle luciferases, aequorins and other photoproteins, for example as described in U.S. Patents 5,221,623, issued June 22, 1989 to Thompson *et al.*, U.S. Patent No. 5,683,888 issued November 4, 1997 to Campbell; U.S. Patent No. 5,674,713 issued September 7 1997 to DeLuca *et al.*, U.S. Patent No. 5,650,289 issued July 22, 1997 to Wood and U.S. Patent No. 5,843,746 issued December 1, 1998 to Tatsumi *et al.* Particularly preferred are bioluminescent proteins isolated from the ostracod *Cypridina* (or *Vargula*) *hilgendorfii*. (Johnson and Shimomura, (1978) Methods Enzymol 57 331-364; Thompson, Nagata & Tsuji (1989) Proc. Natl. Acad. Sci. USA 86, 6567-6571).

22

Beyond the availability of bioluminescent proteins (luciferases) isolated directly from the light organs of beetles, cDNAs encoding luciferases of several beetle species (including, among others, the luciferase of *P. pyralis* (firefly), the four luciferase isozymes of *P. plagiophthalamus* (click beetle), the luciferase of *L. cruciata*(firefly) and the luciferase of *L. lateralis*) (deWet *et al.*, (1987) Molec. Cell. Biol. 7, 725-737; Masuda *et al.*, (1989) Gene 77, 265-270; Wood *et al.*, (1989) Science 244, 700-702; European Patent Application Publication No. 0 353 464) are available. Further, the cDNAs encoding luciferases of any other beetle species, which make bioluminescent proteins, are readily obtainable by the skilled using known techniques (de Wet *et al.* (1986) Meth. Enzymol. 133, 3- 14; Wood *et al.*, (1989) Science 244, 700-702).

Most firefly and click beetle luciferases are ATP- and magnesium dependent and require oxygen for light production. Typically light emission from these enzymes exhibits a rapid burst in intensity followed by a rapid decrease in the first few seconds, followed by a significantly slower sustained light emission. Relatively sustained light output at high rates has been accomplished in these systems by inclusion of coenzyme A, dithiothreitol and other reducing agents that reduce product inhibition and slows inactivation of the luciferase that occurs during catalysis of the light producing reaction, as described in U.S. Patents No. 5,641,641, issued June 24, 1997, and U.S. Patent No. 5,650,289, issued July 22, 1997. Such stable light emitting systems are preferred for use in the present invention.

Particularly preferred bioluminescent proteins are those derived from the ostracod *Cypridina* (or *Vargula*) *hilgendorfii*. The *Cypridina* luciferase (GenBank accession no. U89490) uses no cofactors other than water and oxygen, and its luminescent reaction proceeds optimally at pH 7.2 and physiological salt concentrations, (Shimomura,O., Johnson, F.H. and Saiga, Y. (1961) J. Cell. Comp. Physiol. 58 113-124). By comparison, firefly luciferase has optimal activity at low ionic strength, alkaline pH and reducing conditions, that are typically quite different to those usually found within mammalian cells. Because *Cypridina* luciferase has a turnover number of 1600 min^{-1} and a quantum yield of 0.29, (Shimomura, O. & Johnson, F.H. and Masugi, T. (1969) Science 164 1299-1300; Shimomura, O. & Johnson, F.H. (1970) Photochem. Photobiol. 12 291-295), the *Cypridina* luciferase produces a specific photon flux exceeding that of the optimized firefly system by a factor of at least 50 (Miesenbock and Rothman, (1997) Proc. Natl. Acad. Sci. USA 94 3402-3407).

Naturally Fluorescent Proteins

Another preferred class of embodiments of the reporter moiety includes naturally fluorescent proteins such as the Green Fluorescent Protein (GFP) of *Aequorea victoria* (Tsien, R.Y. (1998) Annu. Rev. Biochem. 67 509-44). Because the entire fluorophore and peptide of a

naturally fluorescent protein can be expressed within intact living cells without the addition of other co-factors or fluorophores, optical probes comprising such proteins as the reporter moiety provide the ability to monitor activities, within defined cell populations, tissues or in an entire transgenic organism. For example, by the use of cell type specific promoters and subcellular

5   targeting motifs, it is possible to selectively target the probe to a discrete location to enable highly spatially defined measurements.

Naturally fluorescent proteins have been isolated and cloned from a number of marine species including the sea pansies *Renilla reniformis, R. kollikeri and R. mullerei* and from the sea pens *Ptilosarcus, Stylatula and Acanthoptilum,* as well as from the Pacific Northwest jellyfish,

10   *Aequorea victoria;* Szent-Gyorgyi *et al.* (SPIE conference 1999); D.C. Prasher *et al.,* (1992) Gene, 111:229-233 and several species of coral (Matz *et al.* (1999). Nature Biotechnology 17 969-973. These proteins are capable of forming a highly fluorescent, intrinsic chromophore through the cyclization and oxidation of internal amino acids within the protein that can be spectrally resolved from weakly fluorescent amino acids such as tryptophan and tyrosine.

15   Additionally naturally fluorescent proteins have also been observed in other organisms, although in most cases these require the addition of some exogenous factor to enable fluorescence development. For example, the cloning and expression of yellow fluorescent protein from *Vibrio fischeri* strain Y-1 has been described by T.O. Baldwin *et al.,* Biochemistry (1990) 29 5509-15. This protein requires flavins as fluorescent co-factors. The cloning of

20   Peridinin-chlorophyll *a* binding protein from the dinoflagellate *Symbiodinium* sp. was described by B.J. Morris *et al.,* (1994) Plant Molecular Biology, 24 673:77. One useful aspect of this protein is that it fluoresces in red. The cloning of phycobiliproteins from marine cyanobacteria such as *Synechococcus,* e.g., phycoerythrin and phycocyanin, is described in S.M. Wilbanks *et al.,* (1993) J. Biol. Chem. 268 1226-35. These proteins require phycobilins as fluorescent co-

25   factors, whose insertion into the proteins involves auxiliary enzymes. The proteins fluoresce at yellow to red wavelengths.

A variety of mutants of the GFP from *Aequorea victoria* have been created that have distinct spectral properties, improved brightness and enhanced expression and folding in mammalian cells compared to the native GFP, (SEQ. ID. NO.: 8), **Table 2.** (*Green Fluorescent*

30   *Proteins,* Chapter 2, pages 19 to 47, edited Sullivan and Kay, Academic Press, U.S. Patent Nos: 5,625,048 to Tsien *et al.,* issued April 29, 1997; 5,777,079 to Tsien *et al.,* issued July 7, 1998; and U.S. Patent No. 5,804,387 to Cormack *et al.,* issued September 8, 1998). In many cases these functional engineered fluorescent proteins have superior spectral properties to wild–type *Aequorea* GFP and are preferred for use as reporter moieties in the present invention.

35

TABLE 2

Aequorea Fluorescent Proteins

Mutations	Common Name	Quantum Yield & Molar Extinction (	Excitation & Emission Max	Relative Fluorescence At 37 °C	Sensitivity To Low pH % max F at pH 6
S65T type					
S65T, S72A, N149K, M153T, I167T	Emerald (SEQ. ID. NO.: 28)		487 509	100	91
F64L, S65T, V163A			488 511	54	43
F64L,S65T	EGFP		488 507	20	57
S65T			489 511	12	56
Y66H type					
F64L, Y66H, Y145F, V163A	P4-3E		384 448	100	N.D.
F64L, Y66H, Y145F			383 447	82	57
Y66H, Y145F	P4-3		382 446	51	64
Y66H	BFP		384 448	15	59
Y66W type					
S65A, Y66W, S72A, N146I, M153T, V163A	W1C			100	82
F64L,S65T, Y66W, N146I, M153T, V163A	W1B			80	71
Y66W, N146I, M153T, V163A	hW7			61	88
Y66W			436 485	N.D.	N.D.
T203Y type					
S65G, S72A, K79R, T203Y	Topaz		514 527	100	14

S65G, V68L, S72A, T203Y	10C		514 527	58	21
S65G,V68L, Q69K, S72A, T203Y	h10C+		516 529	50	54
S65G, S72A, T203H			508 518	12	30
S65G, S72A T203F			512 522	6	28
T203I type					
T203I, S72A, Y145F	Sapphire		395 511	100	90
T203I T202F			395 511	13	80

Non Aequorea, naturally fluorescent proteins, for example Anthozoan fluorescent proteins, and functional engineered homologs thereof, are also suitable for use in the present invention including those shown in **Table 3** below.

5

TABLE 3					
Anthozoa Fluorescent Proteins					
Species	Protein Name	Quantum Yield & Molar Extinction (	Excitation & Emission Max	Relative Brightness	SEQ. ID. NO.:
Anemonia majano	amFP48 6	0.24 40,000	458 486	0.43	SEQ. ID. NO.: 9
Zoanthus sp	zFP506 zFP538	0.63 35,600 0.42 20,200	496, 506 528, 538	1.02 0.38	SEQ. ID. NO.:10 SEQ. ID. NO.:11
Discosoma striata	dsFP483	0.46 23,900	443 483	0.5	SEQ. ID. NO.:12
Discosoma sp "red"	drFP583	0.23 22,500	558 583	0.24	SEQ. ID. NO.:13
Clavularia sp	CFP484	0.48 35,300	456 484	0.77	SEQ. ID. NO.:14

III.   LINKER MOIETIES

Generally linker moieties for measuring a post-translational activity encompass a post-translational recognition motif that contains a residue that, when modified, modulates the coupling of the reporter moiety to the multimerized destabilization domain, thus effecting a change in the stability of the reporter moiety. Typically, for measuring proteases, such linkers contain a single scissile bond (bond that is cleaved within the substrate) for a specific protease and preserve the native function and activity of the reporter moiety and destabilization domains in the intact fusion protein. The design and size of peptide sequences for specific constructs, is dependent upon the application for which the optical probe is to be used. For example, for most applications, the peptide linker separating the reporter moiety and the multimerized destabilization domains will typically be in the range of 5 to 50 amino acids in length, preferably 10 to 25 amino acids in length, or more preferably 10 to 15 amino acids in length. For certain applications, the peptide may be significantly larger, up to and including entire protein domains, for example 50 to 100 amino acids in length. Smaller peptides, in the range of 5 to 50 amino acids may also be used. Typically the protease site may be located at any position within the linker with respect to the reporter moiety and destabilization domains.

In one embodiment the linker comprises a single polypeptide chain that covalently couples the destabilization domains to the reporter moiety. Typically in this embodiment, the linker will comprise a post-translational recognition motif such as a protease recognition motif. Cleavage of the linker by the protease at the cleavage site results in uncoupling of the multimerized destabilization domains from the reporter moiety resulting in a modulation in the stability of the reporter moiety. An important feature of the linker is that it does not contain a protease recognition site for -NH-ubiquitin protein endoproteases that would otherwise result in the post-translational processing of the construct irrespective of the presence or absence of the target post-translational activity. Any cleavage activity capable of hydrolyzing the linker moiety may be assayed with this embodiment of the present invention, provided it does not also cleave the reporter moiety thereby directly modulating its function.

In another aspect of this method, the linker may comprise distinct post-translational recognition motifs and cleavage sites for example, a phosphorylation site and a protease cleavage site, as described in commonly owned U.S. Patent Application No. 09/306,542 filed May 5, 1999. In this case, post-translational modification of the linker results in the modulation of the rate and efficiency of cleavage of the modified linker compared to the non-modified linker. This approach enables the present method to be used to detect a broad range of post translational activities.

In some embodiments, the linker functions to couple a target protein to one or more destabilization domains for the purpose of regulating the concentration of the target protein in

the cell. In this case the linker need not contain a protease cleavage site, and may be significantly smaller, in the order of about 1 to 10 amino acids in length.

In another aspect, the linker may comprise two separate polypeptide chains that are capable of interacting with each other to functionally couple the multimerized destabilization domains to the reporter gene. This approach enables an additional range of post-translational activities to be assayed. In this embodiment, one polypeptide chain is typically covalently coupled to the multimerized destabilization domain, and a separate polypeptide chain is covalently coupled to the reporter moiety. (FIG. 1) Binding of the first polypeptide chain to the second polypeptide chain results in coupling of the destabilization domain to reporter moiety resulting in a modulation of the stability of the reporter moiety. This approach thus enables the identification and detection of protein-protein interactions between defined proteins as well as the ability to detect post-translational modifications that influence these protein-protein interactions.

Examples of suitable interaction domains include protein-protein interaction domains such as SH2, SH3, PDZ, 14-3-3, WW and PTB domains. Other interaction domains are described in for example, the database of interacting proteins available on the web at http://www.doe-mbi.ucla.edu.

To identify and characterize the interaction of two test proteins, the method would typically involve 1) the creation of a first fusion protein comprising the first test protein coupled to the reporter moiety, and a second fusion protein comprising the second test protein coupled to the multimerized destabilization domain construct. 2) The introduction of the test protein fusion proteins alone in to control cells, and in combination into test cells. 3) The measurement of the stability of the reporter moiety in the control cells and test cells. 4) Comparison of the stability of the reporter moiety in the control cells, compared to the stability of the reporter moiety in the test cells. If the cell expressing both test fusion proteins exhibits a reporter moiety with a significantly altered stability (or level of expression) compared to the control cells, then the results indicate that the two proteins do interact under the experimental conditions chosen. Conversely if the stability's of the reporter moieties in the control cells, and in the test cells are the same, then the results indicate that the proteins probably don't interact strongly under the test conditions.

The method also enables the detection and characterization of stimuli (such as receptor stimulation) that cause two proteins to alter their degree of interaction. In this case, a cell line is created that expresses the first and second fusion proteins, as described above, comprising interaction domains that exhibit, or are believed to exhibit post-translational regulated interactions. For example, post-translational modification by phosphorylation of serine or

threonine residues can modulate 14-3-3 domain interactions, tyrosine phosphorylation can influence SH2 domain interactions, the redox state can influence disulfide bond formation. The cell line is then exposed to a test stimulus to determine whether the stimulus regulates the interaction of the two proteins. If the stimulus does regulate the interaction of the two proteins, then this will result in the coupling of the multimerized destabilization domain fusion protein to the reporter moiety fusion protein, subsequently resulting in a modulation of the stability of the reporter moiety in the treated cells, compared to the non-treated cells.

The invention is also readily amenable to identifying new protein-protein interactions. For example, where a first protein is known, but the protein(s) with which it interacts are unknown. In this case, a first fusion protein is made between the first protein and the reporter moiety (or destabilization domain) and cloned into a suitable expression vector. Second, a library of test proteins, for example isolated from a cDNA expression library, is fused in frame to the multimerized destabilization domains (or reporter moiety) and subcloned into a second expression vector. Typically the first fusion protein would be then be introduced into a population of test cells and single clones identified that stably expressed the reporter moiety. The library of test proteins (typically in the form of expression vectors) would be introduced into the clonal cells, stably expressing the first fusion protein. The resulting transformed cells would then be screened to identify cells with altered expression of the reporter moiety fusion compared to the control cells. Suitable clones expressing the reporter moieties with modulated stability, (i.e., reduced levels of the reporter moiety) may then be identified, isolated and characterized, for example by fluorescence activated cell sorting (FACS™). Those library members that display reporter moieties with larger relative changes in expression level may then be identified by the degree to which the stability of the reporter moiety is altered for each library member after exposure to the library of test fusion proteins.

IV.   METHODS OF USE

Introduction of constructs into cells

Typically the constructs of the present invention will be introduced and expressed in target cells via the use of standard molecular biology techniques known in the art. Another approach involves the use of membrane translocating sequences, as described in U.S. patent No. 5,807,746, issued Sept 15 1998 to Lin *et al.* to introduce the protein constructs into cells.

Nucleic acids may also be used to transfect cells with sequences coding for expression of the multimerized destabilization domain, linker and reporter moiety. Generally these will be in the form of an expression vector including expression control sequences

operatively linked to a nucleotide sequence coding for expression of the polypeptide. As used, the term "nucleotide sequence coding for expression of" a polypeptide refers to a sequence that, upon transcription and translation of mRNA, produces the polypeptide. This can include sequences containing, e.g., introns. As used herein, the term "expression control sequences" refers to nucleic acid sequences that regulate the expression of a nucleic acid sequence to which it is operatively linked. Expression control sequences are operatively linked to a nucleic acid sequence when the expression control sequences control and regulate the transcription and, as appropriate, translation of the nucleic acid sequence. Thus, expression control sequences can include appropriate promoters, enhancers, transcription terminators, a start codon (i.e., ATG) in front of a protein-encoding gene, splicing signals for introns, IRES sequences (internal ribosome entry site) maintenance of the correct reading frame of that gene to permit proper translation of the mRNA, and stop codons.

Methods that are well known to those skilled in the art can be used to construct expression vectors containing the multimerized destabilization domain, linker, reporter moiety construct. These methods include in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination/genetic recombination. (See, for example, the techniques described in Maniatis, et al.,(1989) Cold Spring Harbor Laboratory, N.Y.). Many commercially available expression vectors are available from a variety of sources including Clontech (Palo Alto, CA), Stratagene (San Diego, CA) and Invitrogen (San Diego, CA) as well as and many other commercial sources.

A contemplated version of the method is to use inducible controlling nucleotide sequences to produce a sudden increase in the expression of the reporter moiety, linker and multimerized destabilization domain construct e.g., by inducing expression of the construct. Example inducible systems include the tetracycline inducible system first described by Bujard and colleagues (Gossen and Bujard (1992) Proc. Natl. Acad. Sci USA 89 5547-5551, Gossen et al. (1995) Science 268 1766-1769) and described in U.S. Patent No 5,464,758.

Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as E. coli, competent cells that are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method by procedures well known in the art. Alternatively, $MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell or by electroporation.

When the host is a eukaryote, such methods of transfection of DNA as calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection,

electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be co-transfected with DNA sequences encoding the fusion polypeptide of the invention, and a second foreign DNA molecule encoding a selectable phenotype, such as the herpes simplex thymidine kinase gene. Another method is to use an eukaryotic viral vector,

5   such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein. (*Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982). Preferably, an eukaryotic host is utilized as the host cell as described herein.

         The construction of expression vectors and the expression of genes in transfected cells

10   involve the use of molecular cloning techniques also well known in the art. Sambrook *et al.*, (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (most recent Supplement). Nucleic acids used to transfect cells with sequences coding for expression of the polypeptide of interest

15   generally will be in the form of an expression vector including expression control sequences operatively linked to a nucleotide sequence coding for expression of the polypeptide comprising the optical probe.

Assays for post-translational activities

20   In one class of embodiments, the present invention can be used to measure post-translational activities, such as proteolysis, phosphorylation, dephosphorylation, glycosylation, methylation, sulfation, prenylation, disulfide bond formation and ADP-ribsoylation within cells.

         The method generally involves the expression within, or introduction into a cell of a reporter moiety that is functionally coupled to one or more destabilizing domains via a linker.

25   The linker typically contains a recognition motif that is specific for the post-translational activity to be assayed. Modification of the linker by the post-translational activity, results in uncoupling of the reporter moiety from the destabilizing domain resulting in a modulation in the stability of the reporter moiety. The level of activity within a sample is sensed by a measurable change in the level of the reporter moiety, for example by detecting at least one optical property of the

30   reporter moiety, or by detecting at least one optical property of detectable product of the reporter moiety.

         To measure protease activity, it is typically desirable to provide an expression vector in which the expressed fusion gene product comprises a reporter moiety covalently linked to the multimerized destabilization moieties via a single amino acid chain. Thus under these conditions

35   the expressed construct is destabilized until acted upon by the target protease. Upon proteolysis,

31

the cleaved reporter moiety exhibits significantly increased stability, resulting in its steady state accumulation within the cell to a higher level.

The choice of reporter moiety depends in part on the cellular system in which the assays are conducted, and the sensitivity and detection means at hand. For mammalian cells, the -
5 lactamase, -galactosidase, and naturally fluorescent protein based reporter genes provide for intracellular fluorescent measurements, which are preferred. Preferred reporter moieties for luminescent readouts include luciferase and other bioluminescent protein based reporters. In plant studies, preferred reporters include -glucuronidase and luciferase. For transgenic applications in whole animals or intact tissue samples, naturally fluorescent proteins are
10 preferred because the reporter does not require the addition of any substrates or co-factors in order to produce a detectable product. For applications were high sensitivity is required, for example because the target activity has a low turnover number, enzymatic reporter moieties are preferred because they provide enzymatic amplification. That is, each reporter moiety is capable of generating hundreds or thousands of detectable products per minute. By comparison a non
15 enzymatic reporter, such as a naturally fluorescent protein, provides for little signal amplification.

The choice of the multimerized destabilization domain, and the number of copies of the destabilization domain to use are also dependent on the reporter moiety and type of activity being measured. Preferred destabilization domains include, those based on ubiquitin (SEQ. ID.
20 NO.: 2) and mutants and homologs thereof. Particularly preferred are mutants or homologs of ubiquitin (SEQ. ID. NO.: 2) comprising mutations that prevent, or significantly reduce, the cleavage of ubiquitin multimers by -NH-ubiquitin protein endoproteases.

To establish the optimal number of destabilization domains one would generally start by evaluating a construct containing three copies of the destabilization domain. Depending upon
25 the results, one would either increase or decrease the number of copies of destabilization domains. Generally one would increase the number of copies of the destabilization domain if the steady state levels of the non-protease treated samples were too high (too little degradation), and decrease the number of copies of the destabilization domain if the steady state level of the non-protease treated samples were too low (too much degradation). If the target protein was subject
30 to excessive degradation, the steady state level of the target protein may be too low to provide for effective cleavage by the protease, particularly if that protease exhibits a relatively low affinity for that substrate.

An important advantage of the present invention is the ability to titrate the degree of destabilization, and therefore the steady state concentration, of the target protein in the cell.
35 Since the destabilized, unmodified sensor represents the substrate for the target activity, it is

preferable to provide the substrate at a physiologically relevant concentration within the cell while retaining the appropriate turnover characteristics for each individual reporter molecule.

. For assays measuring protease activity, the linker generally comprises a protease recognition motif within its sequence. The protease recognition motif may be placed anywhere within the linker moiety, but is conveniently placed close to the center of the linker unless there are steric, or other reasons, to position the recognition motif at a specific location. Typically, the recognition motif will provide for relatively specific recognition of the sequence by the target protease. In some cases it may be preferable for the linker to contain a second "control" protease site for a known protease for use as a positive control.

The expression vector will normally direct expression of the sensor to the cytosol of the cell, although other cellular compartments, such as the plasma membrane are also practical. Once the expression vector is introduced in a population of cells, the cells are typically screened for reporter moiety expression level in the absence of the target protease. This can be achieved by FACS™, after addition of appropriate substrates for the reporter moieties (if required). While cells may be selected for varying levels of expression of the reporter moiety within the population of cells, observations to date suggest that cells exhibiting somewhat lower levels of reporter moiety are superior to those that initially exhibit high levels of reporter moieties under these conditions. Cells may also be selected via antibiotic resistance to provide for stable cell lines.

Once isolated and characterized, the resulting cell line represents a living sensor for the activation or expression of the target protease that enables the identification and screening of compounds that modulate the activation of the target protease. Importantly these determinations can be completed within the living cell where other issues such as membrane permeability, specificity and toxicity may be directly assessed.

In most cases, it will be preferable to start with a cell line that does not normally express high levels of the active target protease. However if this is not possible, then the initial evaluation of the cell lines may be modified in order to screen for cells initially exhibiting high levels of reporter moiety expression. For example, by using an inhibitor of the reporter moiety to inhibit basal reporter gene activity, (as discussed below). In general any types of cells may be used with the present invention, including animal, plant, insect, yeast and other eukaryotic cells or prokaryotic cells.

In whole cell studies it may be desirable to add an inhibitor of protein synthesis such as cycloheximide in order to reduce the steady state level of the destabilized reporter moiety in the cell immediately prior to the measurement of reporter activity. This approach has the advantage of improving the dynamic range of the assay because in the absence of new protein synthesis,

uncleaved and therefore destabilized reporter moieties are destroyed by targeting to the
proteasome leaving the cleaved and stabilized reporters intact within the cell. (i.e. the
background is reduced). This results in a larger net difference in reporter moiety activity in cells
containing a suitable protease compared with those lacking a suitable protease. Typically for
5   such uses, cycloheximide is added to cell in the range of 10 to 150 g/ml cycloheximide,
preferably 50 to 100 g/ml. Generally cells are pretreated with an appropriate stimulus to
activate the target protease, and then cycloheximide is added one to two hours prior to the
addition of suitable substrates for the reporter moiety.

In another aspect of this method, it sometimes may also be desirable to add an inhibitor
10   of the enzymatic reporter moiety to reduce the activity of the reporter moiety prior to compound
addition in screening applications.   For example, in order to screen for inhibitors of a
constitutively active protease, such inhibitors of reporter activity can be used to eliminate the
pool of cleaved and stabilized reporter prior to adding compound, in effect zeroing out the cells
to begin the experiment.   This approach also has the advantage that the actual concentration of
15   destabilized substrate molecules is not reduced in the cell, so that the protein substrate can be
effectively degraded by the target protease. Example inhibitors include clavulanic acid for the
-lactamase reporter gene (see commonly owned U.S. Patent Application No. 09/067,612 filed
April 28, 1999) and phenylethyl- -D-thiogalactoside for -galactosidase (see Fiering   *et al.*,
(1991) Cytometry 12 291-301). These membrane permeable inhibitors may be added prior to,
20   simultaneously with, or after exposure of the cells to an inhibitor of protein synthesis.

To measure the degree of protein-protein interaction between two defined test proteins, it
is typically desirable to separately couple one protein to one or more destabilization domains,
and the second protein to the reporter moiety, and then express both fusion proteins in a test cell.
This could be achieved for example by transfecting a cell with two compatible expression
25   vectors. In one expression vector, the expressed fusion protein typically comprises a reporter
moiety coupled to the first test protein, and in the second expression vector, the expressed fusion
protein typically comprises the second test protein, coupled to one or more destabilization
domains.

If the first polypeptide fusion protein binds to the second polypeptide fusion protein then
30   the destabilization domain(s) are effectively coupled to the reporter moiety resulting in a
modulation of its stability. Thus the relative degree of destabilization of the reporter moiety is a
direct indicator of the extent to which the proteins physically interact. Typically this can be
accomplished by determining the stability of the reporter moiety in a cell expressing both
proteins compared to a control cell, expressing the reporter moiety fusion protein alone. If the
35   cell expressing both constructs exhibits a reporter moiety with a significantly altered stability

compared to the control cell, the results indicate that the two proteins are interacting when co-expressed within the cell.

The choice and selection of the appropriate reporter moiety and destabilization domain are determined by the same issues of sensitivity and ease of detection discussed above. Preferred
5   reporter moieties include -lactamase and naturally fluorescent proteins. Preferred destabilization domains include those based on ubiquitin (SEQ. ID. NO.: 2), and mutants and functional homologs thereof. Particularly preferred are mutants or homologs of ubiquitin (SEQ. ID. NO.: 2) comprising mutations that prevent, or significantly reduce, the cleavage of ubiquitin multimers by -NH-ubiquitin protein endoproteases.

10   The choice of the number of copies of the destabilization domain is dependent on the affinity of the target interaction to be measured, and the degree of destabilization exerted on the reporter moiety when the proteins are associated. In many cases, the affinity of the interaction will not be known and it will be necessary to evaluate a range of multimerized constructs in order to identify the optimal assay characteristics. Ideally a multimerized construct will be
15   selected in which both the first test protein and the second test protein are present at physiologically relevant concentrations. One way to achieve this result may be to couple both the first test protein and the second test protein with at least one ubiquitin (SEQ. ID. NO.: 2) domain. Under these circumstances both proteins are slowly degraded when separated, but more rapidly degraded when complexed together.

20

Induction and regulation of expression levels of target proteins

In another embodiment, the invention provides for a generalized way of coordinately regulating the cellular concentration of a plurality of target proteins in a cell, or transgenic organism. In this method, the target proteins are operatively coupled to a multimerized
25   destabilization domain via a linker. By varying the number of destabilization domains present in the multimerized destabilization domain, it is possible to titrate the degree of destabilization, and therefore the steady state concentration of the target protein within the cell or transgenic organism. Thus using this approach it is possible to reproducibly vary the relative stoichometery, as well as, the level of expression, of one or more target proteins.

30   In some embodiments the linker may comprise about 1 to 10 amino acids. Typically the linker is non-cleavable by -NH-ubiquitin protein endoproteases.

In one embodiment the linker may contain a non-naturally occurring protease cleavage site (in that cell type), such that cleavage of the linker by the protease results in uncoupling of the target protein from the multimerized destabilization domain hence creating an increase in the
35   stability and concentration of the target protein after protease digestion. In one aspect of this

method, regulation of the activity of the protease can be achieved via regulating the concentration and exposure of the cell to an inhibitor of the protease.

This approach enables the coordinate regulation of the intracellular concentration of a number of target proteins that contain the same protease recognition sites in their linker moieties, simultaneously within a cell. The approach is particularly well suited for the engineering of organisms or cells where multiple proteins need to be induced and expressed in order to create the desired effect, for example for regulating a multi-step metabolic or signal transduction pathway.

In one embodiment the protease is a non-naturally occurring protease in the host cell, which recognizes a relatively rare recognition motif in the linker moiety, for example, including proteases such as Factor Xa (EC 3.4.21.6), Entrokinase (EC 3.4.21.9) and IgA protease (EC 3.4.21.72). Proteases that recognize defined sequences of at least 4, or preferably at least 5 or more preferably about 6 amino acid residues, are generally preferred. Viral proteases, such as a CMV protease or other non-naturally occurring proteases (for that particular cell or organism) are also preferred. If this is the case, then expression of the protease should not significantly impact the cell, and the fusion proteins should not suffer non-specific degradation via the host cells endogenous proteases. Induction or activation of the protease in the cell results in a rapid increase in protease activity within the cell that can cleave the target fusion proteins thereby increasing their stability and steady state concentration in the cell.

V.   SCREENING APPLICATIONS

The present invention is suited for use with systems and methods that utilize automated and integratable workstations for identifying modulators, and chemicals having useful activity. Such systems are described generally in the art (see, U.S. Patent NOs: 4,000,976 to Kramer *et al.* (issued January 4, 1977), 5,104,621 to Pfost *et al.* (issued April 14, 1992), 5,125,748 to Bjornson *et al.* (issued June 30, 1992), 5,139,744 to Kowalski (issued August 18, 1992), 5,206,568 Bjornson *et al.* (issued April 27, 1993), 5,350,564 to Mazza *et al.* (September 27, 1994), 5,589,351 to Harootunian (issued December 31, 1996), and PCT Application Nos: WO 93/20612 to Baxter Deutschland GMBH (published October 14, 1993), WO 96/05488 to McNeil *et al.* (published February 22, 1996), WO 93/13423 to Agong *et al.* (published July 8, 1993) and U.S. Patent No. 5,985,214, issued November 16, 1999.

Typically, such a system includes: A) a storage and retrieval module comprising storage locations for storing a plurality of chemicals in solution in addressable chemical wells, a chemical well retriever and having programmable selection and retrieval of the addressable

36

chemical wells and having a storage capacity for at least 100,000 addressable wells, B) a sample distribution module comprising a liquid handler to aspirate or dispense solutions from selected addressable chemical wells, the chemical distribution module having programmable selection of, and aspiration from, the selected addressable chemical wells and programmable dispensation 5 into selected addressable sample wells (including dispensation into arrays of addressable wells with different densities of addressable wells per centimeter squared) or at locations, preferably pre-selected, on a plate, C) a sample transporter to transport the selected addressable chemical wells to the sample distribution module and optionally having programmable control of transport of the selected addressable chemical wells or locations on a plate (including adaptive routing and 10 parallel processing), D) a reaction module comprising either a reagent dispenser to dispense reagents into the selected addressable sample wells or locations on a plate or a fluorescent detector to detect chemical reactions in the selected addressable sample wells or locations on a plate, and a data processing and integration module.

The storage and retrieval module, the sample distribution module, and the reaction 15 module are integrated and programmably controlled by the data processing and integration module. The storage and retrieval module, the sample distribution module, the sample transporter, the reaction module and the data processing and integration module are operably linked to facilitate rapid processing of the addressable sample wells or locations on a plate. Typically, devices of the invention can process at least 100,000 addressable wells or locations on 20 a plate in 24 hours. This type of system is described in commonly owned U.S. Patent No. 5,985,214, issued November 16, 1999. If desired, each separate module is integrated and programmably controlled to facilitate the rapid processing of liquid samples, as well as being operably linked to facilitate the rapid processing of liquid samples. In one embodiment the system provides for a reaction module that is a fluorescence detector to monitor fluorescence. 25 The fluorescence detector is integrated to other workstations with the data processing and integration module and operably linked with the sample transporter. Preferably, the fluorescence detector is of the type described herein and can be used for epi-fluorescence. Other fluorescence detectors that are compatible with the data processing and integration module and the sample transporter, if operable linkage to the sample transporter is desired can be used as known in the 30 art or developed in the future. For some embodiments of the invention, particularly for plates with 96, 192, 384 and 864 wells per plate, detectors are available for integration into the system. Such detectors are described in U.S. Patent 5,589,351 (Harootunian), U.S. Patent 5,355,215 (Schroeder), and PCT patent application WO 93/13423 (Akong). Alternatively, an entire plate may be "read" using an imager, such as a Molecular Dynamics Fluor-Imager 595 (Sunnyvale, 35 CA). Multi-well platforms having greater than 864 wells, including 3,456 wells, can also be

used in the present invention (see, for example, the PCT Application PCT/US98/11061, filed 6/2/98. These higher density well plates require miniaturized assay volumes that necessitate the use of highly sensitivity assays that do not require washing. The present invention provides such assays as described herein.

The screening methods described herein can be made on cells growing in or deposited on solid surfaces. A common technique is to use a microtiter plate well wherein the fluorescence measurements are made by commercially available fluorescent plate readers. One such method is to use cells in Costar 96 well microtiter plates (flat with a clear bottom) and measure fluorescent signal with CytoFluor multiwell plate reader (Perseptive Biosystems, Inc., MA) using two emission wavelengths to record fluorescent emission ratios. In another embodiment, the system comprises a microvolume liquid handling system that uses electrokinetic forces to control the movement of fluids through channels of the system, for example as described in U.S. patent No., 5,800,690 issued September 1, 1998 to Chow *et al.*, European patent application EP 0 810 438 A2 filed May 5 1997, by Pelc *et al.* and PCT application WO 98/00231 filed 24 June 1997 by Parce *et al.* These systems use "chip" based analysis systems to provide massively parallel miniaturized analysis. Such systems are preferred systems of spectroscopic measurements in some instances that require miniaturized analysis.

A method for identifying a chemical, modulator or a therapeutic

The present invention can also be used for testing a therapeutic for useful therapeutic activity. A therapeutic is identified by contacting a test chemical suspected of having a modulating activity of a biological process or target with a test cell comprising the constructs of the present invention. Typically the cells are located within at least one well of a multi-well platform. The test chemical can be part of a library of test chemicals that is screened for activity, such as biological activity. The library can have individual members that are tested individually or in combination, or the library can be a combination of individual members. Such libraries can have at least two members, preferably greater than about 100 members or greater than about 1,000 members, more preferably greater than about 10,000 members, and most preferably greater than about 100,000 or 1,000,000 members. After appropriate incubation of the sample with the test cell, an inhibitor of protein synthesis may be added and a substrate for the reporter moiety added. At least one optical property (such as fluorescence or absorbance) of the sample is determined and compared to a non-treated control to determine the level of reporter gene expression or activity. If the sample having the test chemical exhibits increased or decreased reporter moiety expression or activity relative to that of the control cell then a candidate modulator has been identified.

The candidate modulator can be further characterized and monitored for structure, potency, toxicology, and pharmacology using well-known methods. The structure of a candidate modulator identified by the invention can be determined or confirmed by methods known in the art, such as mass spectroscopy. For putative modulators stored for extended periods of time, the structure, activity, and potency of the putative modulator can be confirmed.

Depending on the system used to identify a candidate modulator, the candidate modulator will have putative pharmacological activity. For example, if the candidate modulator is found to inhibit a protein tyrosine phosphatase involved, for example in T-cell proliferation *in vitro*, then the candidate modulator would have presumptive pharmacological properties as an immunosuppressant or anti-inflammatory (see, Suthanthiran *et al.*, (1996) Am. J. Kidney Disease, 28 159-172) Such nexuses are known in the art for several disease states, and more are expected to be discovered over time. Based on such nexuses, appropriate confirmatory *in vitro* and *in vivo* models of pharmacological activity, as well as toxicology, can be selected. The assays, and methods of use described herein, enable rapid pharmacological profiling to assess selectivity and specificity, and toxicity. This data can subsequently be used to develop new candidates with improved characteristics.

Bioavailability and Toxicology of Candidate Modulators

Once identified, candidate modulators can be evaluated for bioavailability and toxicological effects using known methods (see, Lu, Basic Toxicology, Fundamentals, Target Organs, and Risk Assessment, Hemisphere Publishing Corp., Washington (1985); U.S. Patent Nos: 5,196,313 to Culbreth (issued March 23, 1993) and U.S. Patent No. 5,567,952 to Benet (issued October 22, 1996). For example, toxicology of a candidate modulator can be established by determining *in vitro* toxicity towards a cell line, such as a mammalian i.e. human, cell line. Candidate modulators can be treated with, for example, tissue extracts, such as preparations of liver, such as microsomal preparations, to determine increased or decreased toxicological properties of the chemical after being metabolized by a whole organism. The results of these types of studies are often predictive of toxicological properties of chemicals in animals, such as mammals, including humans.

The toxicological activity can be measured using reporter genes that are activated during toxicological activity or by cell lysis (see WO 98/13353, published 4/2/98). Preferred reporter genes produce a fluorescent or luminescent translational product (such as, for example, a Green Fluorescent Protein (see, for example, U.S. Patent No. 5,625,048 to Tsien *et al.*, issued 4/29/98; U.S. Patent No. 5,777,079 to Tsien *et al.*, issued 7/7/98; WO 96/23810 to Tsien, published 8/8/96; WO 97/28261, published 8/7/97; PCT/US97/12410, filed 7/16/97; PCT/US97/14595,

filed 8/15/97)) or a translational product that can produce a fluorescent or luminescent product (such as, for example, beta-lactamase (see, for example, U.S. Patent No. 5,741,657 to Tsien, issued 4/21/98, and WO 96/30540, published 10/3/96)), such as an enzymatic degradation product. Cell lysis can be detected in the present invention as a reduction in a fluorescence signal

5   from at least one photon-producing agent within a cell in the presence of at least one photon reducing agent. Such toxicological determinations can be made using prokaryotic or eukaryotic cells, optionally using toxicological profiling, such as described in PCT/US94/00583, filed 1/21/94 (WO 94/17208), German Patent No 69406772.5-08, issued 11/25/97; EPC 0680517, issued 11/12/94; U.S. Patent No. 5,589,337, issued 12/31/96; EPO 651825, issued 1/14/98; and

10   U.S. Patent No. 5,585,232, issued 12/17/96).

Alternatively, or in addition to these *in vitro* studies, the bioavailability and toxicological properties of a candidate modulator in an animal model, such as mice, rats, rabbits, or monkeys, can be determined using established methods (see, Lu, supra (1985); and Creasey, Drug Disposition in Humans, The Basis of Clinical Pharmacology, Oxford University Press, Oxford

15   (1979), Osweiler, Toxicology, Williams and Wilkins, Baltimore, MD (1995), Yang, Toxicology of Chemical Mixtures; Case Studies, Mechanisms, and Novel Approaches, Academic Press, Inc., San Diego, CA (1994), Burrell *et al.*, Toxicology of the Immune System; A Human Approach, Van Nostrand Reinhld, Co. (1997), Niesink *et al.*, Toxicology; Principles and Applications, CRC Press, Boca Raton, FL (1996)). Depending on the toxicity, target organ, tissue, locus, and

20   presumptive mechanism of the candidate modulator, the skilled artisan would not be burdened to determine appropriate doses, LD_{50} values, routes of administration, and regimes that would be appropriate to determine the toxicological properties of the candidate modulator. In addition to animal models, human clinical trials can be performed following established procedures, such as those set forth by the United States Food and Drug Administration (USFDA) or equivalents of

25   other governments. These toxicity studies provide the basis for determining the therapeutic utility of a candidate modulator *in vivo*.

Efficacy of Candidate Modulators

Efficacy of a candidate modulator can be established using several art-recognized

30   methods, such as *in vitro* methods, animal models, or human clinical trials (see, Creasey, supra (1979)). Recognized *in vitro* models exist for several diseases or conditions. For example, the ability of a chemical to extend the life-span of HIV-infected cells *in vitro* is recognized as an acceptable model to identify chemicals expected to be efficacious to treat HIV infection or AIDS (see, Daluge *et al.*, (1995) Antimicro. Agents Chemother. 41 1082-1093). Furthermore, the

35   ability of cyclosporin A (CsA) to prevent proliferation of T-cells *in vitro* has been established as

40

an acceptable model to identify chemicals expected to be efficacious as immunosuppressants (see, Suthanthiran *et al.*, supra, (1996)). For nearly every class of therapeutic, disease, or condition, an acceptable *in vitro* or animal model is available. Such models exist, for example, for gastro-intestinal disorders, cancers, cardiology, neurobiology, and immunology. In addition, these *in vitro* methods can use tissue extracts, such as preparations of liver, such as microsomal preparations, to provide a reliable indication of the effects of metabolism on the candidate modulator. Similarly, acceptable animal models may be used to establish efficacy of chemicals to treat various diseases or conditions. For example, the rabbit knee is an accepted model for testing chemicals for efficacy in treating arthritis (see, Shaw and Lacy, J. (1973) Bone Joint Surg. (Br) 55 197-205. Hydrocortisone, which is approved for use in humans to treat arthritis, is efficacious in this model which confirms the validity of this model (see, McDonough, (1982) Phys. Ther. 62 835-839). When choosing an appropriate model to determine efficacy of a candidate modulator, the skilled artisan can be guided by the state of the art to choose an appropriate model, dose, and route of administration, regime, and endpoint and as such would not be unduly burdened.

In addition to animal models, human clinical trials can be used to determine the efficacy of a candidate modulator in humans. The USFDA, or equivalent governmental agencies, have established procedures for such studies (see, www.fda.gov).

Selectivity of Candidate Modulators

The *in vitro* and *in vivo* methods described above also establish the selectivity of a candidate modulator. It is recognized that chemicals can modulate a wide variety of biological processes or be selective. Panels of cells, each containing constructs with varying specificity, based on the present invention, can be used to determine the specificity of the candidate modulator. Selective modulators are preferable because they have fewer side effects in the clinical setting. The selectivity of a candidate modulator can be established *in vitro* by testing the toxicity and effect of a candidate modulator on a plurality of cell lines that exhibit a variety of cellular pathways and sensitivities. The data obtained from these *in vitro* toxicity studies can be extended into *in vivo* animal model studies, including human clinical trials, to determine toxicity, efficacy, and selectivity of the candidate modulator suing art-recognized methods.

An identified chemical, modulator, or therapeutic and compositions

The invention includes compositions, such as novel chemicals, and therapeutics identified by at least one method of the present invention as having activity by the operation of methods, systems or components described herein. Novel chemicals, as used herein, do not

include chemicals already publicly known in the art as of the filing date of this application. Typically, a chemical would be identified as having activity from using the invention and then its structure revealed from a proprietary database of chemical structures or determined using analytical techniques such as mass spectroscopy.

5   One embodiment of the invention is a chemical with useful activity, comprising a chemical identified by the method described above. Such compositions include small organic molecules, nucleic acids, peptides and other molecules readily synthesized by techniques available in the art and developed in the future. For example, the following combinatorial compounds are suitable for screening: peptoids (PCT Publication No. WO 91/19735, 26 Dec. 1991), encoded peptides (PCT Publication No. WO 93/20242, 14 Oct. 1993), random bio-oligomers (PCT Publication WO 92/00091, 9 Jan. 1992), benzodiazepines (U.S. Patent No. 5,288,514), diversomeres such as hydantoins, benzodiazepines and dipeptides (Hobbs DeWitt, S. et al., (1993) Proc. Nat. Acad. Sci. USA 90 6909-6913), vinylogous polypeptides (Hagihara et al., (1992) J. Amer. Chem. Soc. 114 6568), nonpeptidal peptidomimetics with a Beta-D-Glucose scaffolding (Hirschmann, R. et al., (1992) J. Amer. Chem. Soc. 114 9217-9218), analogous organic syntheses of small compound libraries (Chen, C. et al., (1994) J. Amer. Chem. Soc. 116 2661), oligocarbamates (Cho, C.Y. et al., (1993) Science 261: 1303), and/or peptidyl phosphonates (Campbell, D.A. et al., (1994) J. Org. Chem. 59 658). See, generally, Gordon, E. M. et al., (1994). J. Med Chem. 37 1385. The contents of all of the aforementioned publications 20 are incorporated herein by reference.

The present invention also encompasses the identified compositions in a pharmaceutical composition comprising a pharmaceutically acceptable carrier prepared for storage and subsequent administration, which have a pharmaceutically effective amount of the products disclosed above in a pharmaceutically acceptable carrier or diluent. Acceptable carriers or 25 diluents for therapeutic use are well known in the pharmaceutical art, and are described, for example, in Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.R. Gennaro edit. 1985). Preservatives, stabilizers, dyes and even flavoring agents may be provided in the pharmaceutical composition. For example, sodium benzoate, acsorbic acid and esters of p-hydroxybenzoic acid may be added as preservatives. In addition, antioxidants and suspending 30 agents may be used.

The compositions of the present invention may be formulated and used as tablets, capsules or elixirs for oral administration; suppositories for rectal administration; sterile solutions, suspensions for injectable administration; and the like. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or 35 suspension in liquid prior to injection, or as emulsions. Suitable excipients are, for example,

water, saline, dextrose, mannitol, lactose, lecithin, albumin, sodium glutamate, cysteine hydrochloride, and the like. In addition, if desired, the injectable pharmaceutical compositions may contain minor amounts of nontoxic auxiliary substances, such as wetting agents, pH buffering agents, and the like. If desired, absorption enhancing preparations (e.g., liposomes) may be utilized.

The pharmaceutically effective amount of the composition required as a dose will depend on the route of administration, the type of animal being treated, and the physical characteristics of the specific animal under consideration. The dose can be tailored to achieve a desired effect, but will depend on such factors as weight, diet, concurrent medication and other factors which those skilled in the medical arts will recognize. In practicing the methods of the invention, the products or compositions can be used alone or in combination with one another or in combination with other therapeutic or diagnostic agents. These products can be utilized *in vivo*, ordinarily in a mammal, preferably in a human, or *in vitro*. In employing them *in vivo*, the products or compositions can be administered to the mammal in a variety of ways, including parenterally, intravenously, subcutaneously, intramuscularly, colonically, rectally, nasally or intraperitoneally, employing a variety of dosage forms. Such methods may also be applied to testing chemical activity *in vivo*.

As will be readily apparent to one skilled in the art, the useful *in vivo* dosage to be administered and the particular mode of administration will vary depending upon the age, weight and mammalian species treated, the particular compounds employed, and the specific use for which these compounds are employed. The determination of effective dosage levels, that is the dosage levels necessary to achieve the desired result, can be accomplished by one skilled in the art using routine pharmacological methods. Typically, human clinical applications of products are commenced at lower dosage levels, with dosage level being increased until the desired effect is achieved. Alternatively, acceptable *in vitro* studies can be used to establish useful doses and routes of administration of the compositions identified by the present methods using established pharmacological methods.

In non-human animal studies, applications of potential products are commenced at higher dosage levels, with dosage being decreased until the desired effect is no longer achieved or adverse side effects disappear. The dosage for the products of the present invention can range broadly depending upon the desired affects and the therapeutic indication. Typically, dosages may be between about 10 mg/kg and 100 mg/kg body weight, and preferably between about 100 μg/kg and 10 mg/kg body weight. Administration is preferably oral on a daily basis.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g., Fingl *et al.*, in The Pharmacological Basis

43

of Therapeutics, 1975). It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administrated dose in the management of the disorder of interest will vary with the severity of the condition to be treated and to the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above may be used in veterinary medicine.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in Remington's Pharmaceutical Sciences, 18th Ed., Mack Publishing Co., Easton, PA (1990). Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For such transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art. Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular, those formulated as solutions, may be administered parenterally, such as by intravenous injection. The compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external micro-environment and, because liposomes fuse with cell membranes, are efficiently delivered into the cell cytoplasm.

Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions. The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, for example, by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses. For this purpose, concentrated sugar solutions may be used, which

may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses. Such formulations can be made using methods known in the art (see, for example, U.S. Patent Nos. 5,733,888 (injectable compositions); 5,726,181 (poorly water soluble compounds); 5,707,641 (therapeutically active proteins or peptides); 5,667,809 (lipophilic agents); 5,576,012 (solubilizing polymeric agents); 5,707,615 (anti-viral formulations); 5,683,676 (particulate medicaments); 5,654,286 (topical formulations); 5,688,529 (oral suspensions); 5,445,829 (extended release formulations); 5,653,987 (liquid formulations); 5,641,515 (controlled release formulations) and 5,601,845 (spheroid formulations).

VII. TRANSGENIC ANIMALS

In another embodiment, the invention provides a transgenic non-human organism that expresses a nucleic acid sequence that encodes a target protein, (such as a reporter moiety, enzyme or structural protein) functionally coupled to one or more destabilization domains by a linker. Because such constructs can be expressed within intact living cells, with preset degrees of stability, the invention provides the ability to regulate the expression level of the target protein, or to monitor post translational activities within defined cell populations, tissues or in an entire transgenic organism.

In one embodiment the approach may be used to regulate the expression level of an enzyme or group of enzymes involved in a particular signal transduction, disease, or metabolic pathway. Such methods may be useful, for example, for creating transgenic model animals for certain disease states, or for modulating the intracellular concentration of enzymatic intermediates though the manipulation of the expression levels of the enzymes involved. For example, to increase the intracellular concentration of an intermediate one could increase the concentration of the enzyme(s) involved in the synthesis of the intermediate, and / or decrease the concentration of the enzyme(s) involved in degradation of the intermediate. Typically the approach would require the replacement of the native enzymes with fusion proteins of the enzymes with the multimerized destabilization domains of the present invention. For target proteins in which the desired concentration was relatively high, one would select fusion proteins with relatively few (i.e. one or two), or even no, (zero) copies of the destabilization domain. For target proteins for which a relatively low intracellular concentration was desired, one would select fusion proteins with relatively more copies of the destabilization domain (i.e. three or more).

In another embodiment, the approach can be used to identify where in specific tissues a particular activity is located, for example, by expression of a reporter moiety coupled to the multimerized destabilization domain via a linker comprising recognition and cleavage motifs for that activity, in the organism. Typically the linker would comprise a single polypeptide chain that covalently couples the destabilization domains to the reporter moiety. Typically in this embodiment, the linker will comprise a post-translational recognition motif such as a protease recognition motif. Cleavage of the linker by the protease at the cleavage site results in uncoupling of the multimerized destabilization domains from the reporter moiety resulting in a modulation in the stability of the reporter moiety, thereby resulting in an accumulation of reporter moiety in cells or tissues that exhibit protease activity.

Such non-human organisms include vertebrates such as rodents, fish such as Zebrafish, non-human primates and reptiles as well as invertebrates. Preferred non-human organisms are selected from the rodent family including rat and mouse, most preferably mouse. The transgenic non-human organisms of the invention are produced by introducing transgenes into the germline of the non-human organism. Embryonic target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the organism and stage of development of the embryonic target cell. In vertebrates, the zygote is the best target for microinjection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter, which allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82 4438-4442,). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

A transgenic organism can be produced by cross-breeding two chimeric organisms which include exogenous genetic material within cells used in reproduction. Twenty-five percent of the resulting offspring will be transgenic *i.e.*, organisms that include the exogenous genetic material within all of their cells in both alleles. 50% of the resulting organisms will include the exogenous genetic material within one allele and 25% will include no exogenous genetic material.

Retroviral infection can also be used to introduce transgene into a non-human organism. In vertebrates, the developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retro viral infection (Jaenich, R., (1976) Proc.

Natl. Acad. Sci USA 73 1260-1264,). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan, et al. (1986) in Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). The viral vector system used to introduce the transgene is typically a replication-defective retro virus

5   carrying the transgene (Jahner, et al., (1985) Proc. Natl. Acad. Sci. USA 82 6927-6931; Van der Putten, et al., (1985) Proc. Natl. Acad. Sci USA 82 6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, supra; Stewart, et al., (1987) EMBO J. 6 383-388).

          Alternatively, infection can be performed at a later stage. Virus or virus-producing

10   cells can be injected into the blastocoele (D. Jahner et al., (1982) Nature 298 623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells that formed the transgenic nonhuman animal. Further, the founder may contain various retro viral insertions of the transgene at different positions in the genome that generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line, albeit

15   with low efficiency, by intrauterine retro viral infection of the midgestation embryo (D. Jahner et al., supra). A third type of target cell for transgene introduction for vertebrates is the embryonic stem cell (ES). ES cells are obtained from pre-implantation embryos cultured in vitro and fused with embryos (M. J. Evans et al. (1981) Nature 292 154-156; M.O. Bradley et al., (1984) Nature 309 255-258; Gossler, et al., (1986) Proc. Natl. Acad. Sci USA 83 9065-9069; and Robertson et

20   al., (1986) Nature 322 445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retro virus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a nonhuman animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. (For review see Jaenisch, R., (1988) Science 240 1468-1474).

25

VIII TRANSGENIC PLANTS

In another embodiment, the invention provides a transgenic plant that expresses a nucleic acid sequence that encodes a target protein, (such as a reporter moiety, enzyme or structural protein) functionally coupled to a multimerized destabilization domain by a linker. Because such constructs can be specifically expressed, both spatially and temporally, within intact living cells, the invention provides the ability to regulate the expression level of the target protein, within defined cell populations, tissues, or in the entire transgenic plant.

In one embodiment the approach may be used to regulate the expression level of an enzyme or group of enzymes involved in a particular signal transduction, developmental or metabolic pathway. Such methods may be useful for creating transgenic plants with improved disease resistance or other favorable traits. More particularly, plants can be genetically engineered to express various phenotypes of agronomic interest, for example by allowing for the regulated expression of agronomically important genes. Given potential concerns about the safety of transgenic plants, the ability to reduce or eliminate the expression of certain resistance genes prior to harvesting and human consumption is of particular interest. Examples of the types of genes that could be manipulated using the methods described herein, include disease resistance genes, herbicide resistance genes and genes that improve plant traits, including those shown in **Table 4**, below.

TABLE 4		
I. Disease Resistance Genes		
Gene or Gene Product	**Function**	**Reference**
Tomato Cf-9 gene	Resistance to Cladosporium fulvum	Jones et al., Science 266 789 (1994)
Tomato Pto gene	Resistance to Pseudomonassy.ringae	Martin et al., Science 262: 1432 (1993)
Arabidopsis RSP2 gene	Resistance to Pseudomonas syringae	Mindrinos et al., Cell 78: 1089 (1994)
Bacillus thuringiensis protein	Insect resistance	Geiser et al., Gene 48: 109 (1986),
Streptomyces nitrospoeus -amylase inhibitor	Inhibition of amylase activity.	Sumitani et al., Biosci. Biotech. Biochem. 57 1243 (1993)
Expression of insect-specific hormones or pheromones such as an ecdysteroid and juvenile hormone	Disruption of insect development	Hammock et al., Nature 344: 458 (1990)

Expression insect-specific scorpion venom	Insect resistance	Pang et al.,Gene 116: 165 (1992)
Altered expression of metabolic enzymes	Expression of enzymes responsible for the formation of non protein molecules with insecticidal activity	
Altered expression of signal transduction enzymes	Expression of enzymes responsible for the post-translational modification of biologically active molecules	See PCT application WO 93/02197, Botella et al., Plant Molec. Biol. 24: 757 (1994),
Expression of synthetic antimicrobial peptides, such as peptide derivatives of Tachyplesin	Improved disease resistance	
Altered expression of Ion channels, blockers or permeases such as cecropin-3 lytic peptide	Improved resistance to Pseudomonas solanacearum.	Jaynes et al., Plant Sci. 89: 43 (1993),
Expression of viral coat proteins or viral-invasive proteins or toxins.	Improved viral resistance to alfalfa mosaic virus, cucumber mosaic virus, tobacco streak virus, potato virus X, potato virus Y, tobacco etch virus, tobacco rattle virus and tobacco mosaic virus	See Beachy et al., Ann. Rev. Phytopathol. 28: 451 (1990).
Expression of insect-specific antibody or immunotoxins	Improved resistance to insects	Taylor et al., Abstract #497, SEVENTH INT'L SYMPOSIUM ON MOLECULAR PLANT-MICROBE INTERACTIONS (1994)
Expression of virus-specific antibodies.	Improved resistance to viruses	Tavladoraki et al., Nature 366: 469 (1993)
Expression of developmental -arrestive proteins or gene products, as endo a1,4-D-polygalacturonase, or expression of barley ribosome- inactivating gene	Increased resistance to pathogens or parasites	See Lamb et al., Biol Technology ;Q: 1436 (1992). Logemann et al.,BiolTechnology.10: 30 (1992)
II. Herbicide Resistance Genes		
Expression of mutant ALS and AHAS enzymes	Inhibition of the growing point or meristem, increasing resistance to herbicides	Lee et al., EMBO J. 7: 1241 (1988), and Miki et al., Theor. Appl. Genet. 8 : 449 (1990),
Expression of mutant EPSP synthase and aroA	Resistance to glyphosate and other phosphono	U.S. patent No. 4,940,835 to Shah et al., U.S. patent

genes,	compounds such as glufosinate	No. 4,769,061 to Comai. European patent application No. 0 333 033 to Kumada et al. and U.S. patent No. 4,975,374 to Goodman et al.
III. Genes That Confer Or Contribute To A Value-added Trait		
Expression of antisense gene of stearoyl-ACP desaturase	Improved fatty acid composition	Knultzon et al., Proc. Natl. Acad. Sci.USA 89: 2624 (1992).
Expression of phytic acid degrading enzymes	Improved free phosphate composition	Van Hartingsveldt et al., Gene 127: 87 (1993)
Expression of fructosyltransferase, levansucrase, or invertase genes	Improved carbohydrate composition	See Shiroza et al., J. Bacteriol. 170: 810 (1988), Steinmetz et al., Mol. Gen. Genet. 200 220 (1985), Elliot et al., Plant Molec. Biol. 21 515 (1993)

In another embodiment, the approach can be used to specifically identify where in specific tissues a particular activity is expressed, for example by expression of the protease sensor in specific plant tissues.

5   Transgenic plants may be produced by any one of a number of methods of plant transformation and regeneration.   Numerous methods for plant transformation have been developed, including biological and physical, plant transformation protocols. See, for example, Miki et al., "Procedures for Introducing Foreign DNA into Plants" in Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc. , Boca

10   Raton, 1993) pages 67-88. In addition, expression vectors and in vitro culture methods for plant cell or tissue transformation and regeneration of plants are available. See, for example, Gruber et al., "Vectors for Plant Transformation" in Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pages 89-119.

15   The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of Agrobacterium. See, for example, Horsch et al., (1985) Science 227 1229. A. tumefaciens and A. rhizogenes are plant pathogenic soil bacteria which genetically transform plant cells. The Ti

and Ri plasmids of A. tumefaciens and A. rhizogenes, respectively, carry genes responsible for

20   genetic transformation of the plant See, for example, Kado,

C.I., Crit. Rev. Plant. Sci. 10: 1 (1991). Descriptions of Agrobacterium vector systems and methods for Agrobacterium-mediated gene transfer are provided by Gruber et al., supra, Miki et al., supra, and Moloney et al., (1989) Plant Cell Reports 8 238.

5 Despite the fact the host range for Agrobacterium mediated transformation is broad, some major cereal crop species and gymnosperms have generally been recalcitrant to this mode of gene transfer, even though some success has recently been achieved in rice. Hiei et al., (1994) The Plant Journal 6 271-282. Several methods of plant transformation, collectively referred to as direct gene transfer, have been developed as an alternative to Agrobacterium-mediated transformation.

10 A generally applicable method of plant transformation is microprojectile-mediated transformation wherein DNA is carried on the surface of microprojectiles
measuring 1 to 4 Am. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate plant cell walls and membranes. Sanford et al., (1987), Part. Sci. Technol. 5 27,
15 Sanford, J.C., (1988) Trends Biotech. 6 299, Sanford, J.C., (1990) Physiol. Plant 79 206, Klein et al., (1992) Biotechnology 10 268.

Another method for physical delivery of DNA to plants is sonication of target cells. Zhang et al., (1991) BioTechnology 9 996. Alternatively, liposome or spheroplast fusion have been used to introduce expression vectors into plants. Deshayes et al., (1985) EMBO J., 4 2731,
20 Christou et al., (1987) Proc Natl. Acad. Sci. U.S.A. 84 3962. Direct uptake- of DNA into protoplasts using $CaCl_2$ precipitation, polyvinyl alcohol or poly-Lornithine have also been reported. Hain et al., (1985) Mol.Gen. Genet. 199 161 and Draper et al., (1982) Plant Cell Physiol. 23 451. Electroporation of protoplasts and whole cells and tissues have also been described. Donn et al., In Abstracts of VIIth International Congress on Plant Cell and Tissue
25 Culture IAPTC, A2-38, p 53 (1990) ; D'Halluin et al., (1992) Plant Cell 4 1495-1505 and Spencer et al., (1994) Plant Mol. Biol. 24 51-61.

A preferred method is microprojectile-mediated bombardment of immature embryos. The embryos can be bombarded on the embryo axis side to target the meristem at a very early stage of development or bombarded on the scutellar side to target cells that typically form callus and
30 somatic embryos. Targeting of the scutellum using projectile bombardment is well known to those in the art of cereal tissue culture. Klein et al., (1988) BioTechnol. , 6 559-563; Sautter et al., BiolTechnol., 9 1080-1085 (1991) ; Chibbar et al., (1991) Genome, 34 435-460. The scutellar origin of regenerable callus from cereals is well known. Green et al., (1975) Crop Sci., 15 417-421; Lu et al., (1982)TAG 62 109-112; and Thomas and Scott, (1985) J. Plant Physiol.
35 121 159-169 - Targeting the scutellum and then using chemical selection to recover transgenic

plants is well established in cereals. D/Halluin et al., Plant Cell 4: 1495-1505 (1992) ; Perl et al., MGG 235: 279-284 (1992); Cristou et al.,

BioITechnol. 9: 957-962 (1991). This literature reports DNA targeting of the scutellum and recovery of transgenic callus, plants and progeny based on a chemical selection regime. None of these references teach successful plant transformation wherein transformed cells are visualized with a screenable marker such as GUS.

A preferred transformation method involves bombardment of the scutellar surface of immature embryos to introduce the expression cassette with the gene for a bioluminescent protein, such as *Aequorea victoria* GFP (See PCT publication WO 97/41228 to Gordon-Kamm et al., incorporated herein by reference). Embryos can be pretreated for 1 to 48 hours with a high osmoticum medium or left on a highosmoticum medium for 24-48 hours after bombardment to improve cell survival and transformation frequencies. Immature embryos are then cultured on typical callusinducing medium with no selective agent. At each subculture transfer, i.e., every two weeks, the culture is monitored using UV-blue light for GFP fluorescence. Fluorescing calli are separated from non-fluorescing callus, and grown to the point where plants can be regenerated through standard media progressions.

Plants can be manipulated, for example, by removal of the apical meristem, to stimulate axillary or secondary buds which can exhibit larger transgenic sectors relative to the primary shoot. Flowers above transgenic shoots are pollinated and the progeny are analyzed for transgene presence and expression. A variety of starting explants can regenerate shoots in sunflower, and thus represent alternative targets for GFP-encoding DNA delivery and transmission to progeny. These include the seedling meristem (as above), also the seedling hypocotyl, the mature cotyledon, the immature cotyledon, zygotic immature embryos, somaticembryos, and primary leaflets. See for example, respectively, Greco et al., (1984) Plant Sci. Lett. 36 73-77; Krauter et al., (1991) Helia 14 117-122; Power (1987) Am. J. Bot. 74 497503; Krauter et al., (1991) Theor. Appl. Genet. a2: 521525; Finer, (1987) Plant Cell Rep. J: 372-374, and Greco et al., (1984) Plant Sci. Lett. 36 73-77.

EXAMPLES

Example 1 Generation of multimerized destabilization domains

The cDNA encoding human ubiquitin was isolated from a human genomic DNA preparation obtained from Jurkat cells by polymerase chain reaction (PCR) using the PCR

primers Ubi5 (SEQ. ID. NO. 15) and Ubi3 (SEQ. ID. NO. 16) and cloned into pBluescript II vector (Stratagene). The C-terminal residue of ubiquitin was altered from glycine to valine by site-directed mutagenesis (Kunkel) in order to generate a mutant form of ubiquitin that cannot be cleaved by cellular -NH-ubiquitin endopeptidases. This mutant is hereafter referred to as

5 ubiquitinG76V (SEQ. ID. NO. 17). The ubiquitinG76V (SEQ. ID. NO. 17) mutant was then amplified by PCR using the oligonucleotide primers Ub5′ (SEQ. ID. NO. 18) and Ub3′, (SEQ. ID. NO. 19). These primers introduce a Bgl II restriction site at the 5 end of the coding sequence and a BamH I site at the 3 end of the coding sequence. The PCR fragment from the reaction was digested with Bgl II and BamH I and ligated into BamH I-digested pBluescript II

10 vector. This plasmid was then digested with Bgl II and BamH I and the ubiquitinG76V (SEQ. ID. NO. 17) containing fragment was isolated and ligated to generate multimerized ubiquitinG76V domains. The ubiquitinG76V multimers were digested with Bgl II and BamH I to ensure that the individual ubiquitinG76V domains (SEQ. ID. NO. 17) were in the correct orientation.   The digested ubiquitinG76V multimers were separated by agarose gel

15 electrophoresis and multimers of the appropriate sizes were isolated and cloned into BamH I-digested pBluescript II. The ubiquitinG76V multimers were then excised using BamH I and Hind III and subcloned to generate a series of plasmids containing in frame fusions of from one to four copies of ubiquitinG76V (SEQ. ID. NO. 17) fused to the reporter moiety or protein of interest. These constructs are referred to as 1XUb (one copy of ubiquitinG76V (SEQ. ID. NO.

20 17)), 2XUb (two copies of ubiquitinG76V (SEQ. ID. NO. 17)), 3XUb (three copies of ubiquitinG76V (SEQ. ID. NO. 17)) and 4XUb (four copies of ubiquitinG76V (SEQ. ID. NO. 17)).

Example 2. Creation of multimerized destabilization domain--lactamase fusion proteins

25 The gene encoding the E. coli TEM-1 -lactamase was isolated from the plasmid pBluescript (Stratagene) by polymerase chain reaction (PCR) amplification using the PCR primers BLA5 (SEQ. ID. NO. 20) and ABSC107, (SEQ. ID. NO. 21) resulting in the deletion of the signal sequence and introduction of a BamH I restriction site and the amino acids below at the 5 end of the coding sequence.

30
            BamH1
            H G S G A W L H P E T L V K V K

Amino acids in bold represent original -lactamase coding sequence, underlined amino acids represent the BamH I restriction site. An Xba I site was inserted at the 3 end of the coding

35 sequence. The PCR fragments from these reactions were digested with BamH I and Xba I and ligated into pcDNA3 (Invitrogen) via the same sites. The resulting construct, pcDNA3-Bla

(SEQ. ID. NO. 22), was then used to create in-frame fusions with the multimerized ubiquitinG76V constructs above. This was achieved by digesting the multimerized ubiquitinG76V constructs with the restriction enzymes BamH I and Hind III, and then ligating them via the same sites into the pcDNA3-Bla construct. These constructs were named pcDNA3-

5   1XUb-Bla (SEQ. ID. NO. 23), pcDNA3-2XUb-Bla (SEQ. ID. NO. 24), pcDNA3-3XUb-Bla (SEQ. ID. NO. 25), pcDNA3-4XUb-Bla (SEQ. ID. NO. 26). To produce the wild-type - lactamase protein, we used a construct that contains one copy of wild-type (cleavable) ubiquitin (SEQ. ID. NO. 2) fused to the -lactamase coding region in the pcDNA3 vector; this plasmid is referred to as pcDNA3-Ub-Met-Bla (SEQ. ID. NO. 27). Upon synthesis of the Ub-Met-Bla

10   fusion protein, ubiquitin isopeptidases efficiently cleave off the N-terminal ubiquitin (SEQ. ID. NO. 2) precisely after glycine-76, generating the wild-type -lactamase protein with methionine at its N-terminus.

Example 3. Creation of multimerized destabilization domain-Naturally Fluorescent Protein

15   **fusions**

The gene encoding the GFP mutant Emerald (S65T, S72A, N149K, M153T, I167T) (SEQ. ID. NO. 28) was amplified by PCR using the oligonucleotides GFP5(SEQ. ID. NO. 29) and GFP3 (SEQ. ID. NO. 30). The resulting PCR product had a BamH I restriction site at the 5 end of the coding sequence and a Xba I site at the 3 end of the coding sequence. The PCR

20   fragment from this reaction was digested with BamH I and Xba I and ligated into pcDNA3 via the same sites. The resulting construct, pcDNA3-GFP was then used to create in-frame fusions with the multimerized ubiquitinG76V constructs described above. This was achieved by digesting the pcDNA3-1-4XUb-Bla constructs (SEQ. ID. NOs. 23 to 26) with the restriction enzymes BamH I and Hind III, and then ligating the fragment encoding the various multiUb

25   destabilization domains via the same sites into the pcDNA3-GFP construct. These constructs were named pcDNA3-1XUb-GFP (SEQ. ID. NO. 31), pcDNA3-2XUb-GFP (SEQ. ID. NO. 32), pcDNA3-3XUb-GFP (SEQ. ID. NO. 33), pcDNA3-4XUb-GFP (SEQ. ID. NO. 34).

Example 4. Creation of multimerized destabilization domain-Naturally Occurring

30   **Mammalian Protein fusions**

Fusions between multimerized uncleavable ubiquitinG76V (SEQ. ID. NO. 17) and caspase-3 were constructed to further investigate the relationship between the degree of destabilization exerted by varying the number of copies of the destabilization domain with

35   different target proteins.

The caspase-3 cDNA (SEQ. ID. NO. 35) was amplified by PCR using the primers C35 (SEQ. ID. NO. 36) and C33(SEQ. ID. NO. 37) to add BamH I sites at the ends of the caspase-3 cDNA. The amplified caspase-3 cDNA was digested with BamH I then cloned into BamH I-digested pcDNA3-1-4XUb-Bla plasmids (SEQ. ID. NOs. 23 to 26), to create fusions of the

5   different multiubiquitin destabilization domains to a caspase-3- -lactamase fusion. The -lactamase coding region was then removed from these plasmids by digesting to completion with Xba I followed by a partial digest with BamH I. The digests were separated by agarose gel electrophoresis and the correct size DNA band was purified from the gel. The ends of the digested plasmid were blunted with the Klenow fragment of DNA polymerase and the plasmid

10 recircularized by ligation. The resulting plasmids contained an in-frame fusion of the ubiquitinG76V destabilization domain (with from one to four copies of ubiquitinG76V (SEQ. ID. NO. 17)) to the caspase-3 coding region. These plasmids were designated pcDNA3-1-4XUb-C3 (SEQ. ID. NO. 38 to 41). To produce the wild-type caspase-3 protein, the caspase-3 cDNA was amplified by PCR with primers C35Met (SEQ. ID. NO. 42) and C33(SEQ. ID. NO.

15 43) and cloned directly into pcDNA3-Ub-Met-Bla (SEQ. ID. NO. 27). The resulting plasmid was then digested with BamH I and Xba I and recircularized as described above to create the wild-type caspase-3 control construct; this plasmid was designated as pcDNA3-Ub-Met-C3 (SEQ. ID. NO. 44). Upon synthesis of the Ub-Met-caspase-3 fusion protein, ubiquitin isopeptidases efficiently cleave off the N-terminal ubiquitin precisely after glycine-76,

20 generating the wild-type caspase-3 protein with methionine at its N-terminus (data not shown).

Example 5. Characterization of multimerized destabilization domain– -lactamase fusion proteins *in vitro*.

^{35}S-Labeled multimerized destabilization domain– -lactamase fusion protein molecules

25 were produced using a coupled *in vitro* transcription/translation system based on a rabbit reticulocyte lysate (TNT T7 Quick; Promega). Constructs containing from one to four copies of the destabilization domain (pcDNA3-1-4XUb-Bla (SEQ. ID. NOs. 23 to 26) from Example 2) were incubated in the TNT lysate essentially as described in the manufacturer's directions in the presence of 0.25 mCi/ml ^{35}S-methionine (10 mCi/ml, 1175 Ci/mmol; New England Nuclear) to

30 generate ^{35}S-labeled fusion proteins.

To determine the half life of the constructs, 1 l samples of the synthesis reactions were incubated at 37°C in 9 l of chase extract (crude rabbit reticulocyte lysate (Promega) supplemented with 100 g/ml cycloheximide, 1 mM ATP, 20 mM phosphocreatine, 2.5 mM MgCl₂, 5 g/ml creatine kinase, 200 g/ml ubiquitin, and 50 M methionine). The rabbit

35 reticulocyte lysate system contains all of the components necessary for efficient recognition and

degradation of proteins by the ubiquitin-proteasome pathway. Samples were removed at 0, 5, 10, 20, 30, 45 and 60 minutes of reaction and analyzed by polyacrylamide gel electrophoresis (SDS-PAGE). The gels were treated with Amplify (Amersham) and the labeled species detected by autoradiography. This analysis showed that wild-type -lactamase was stable over the 1 hour
5   chase period while the ubiquitinG76V- -lactamase fusions were considerably less stable (FIG. 2A). In particular, the 1XUb-Bla fusions were modestly destabilized ($t_{1/2}$ ~20 min) and - lactamase fusions containing 2, 3 or 4 copies of ubiquitinG76V (SEQ. ID. NO. 17) were strongly destabilized ($t_{1/2}$ <5 min). In addition, the degradation of the 2XUb-Bla fusion was slightly slower than the degradation of -lactamase fusions containing 3 or 4 copies of ubiquitinG76V
10 (SEQ. ID. NO. 17) (FIG. 2A).

In order to test whether the degradation of multiUb-Bla fusions *in vitro* is dependent on the proteasome, TNT synthesis reactions were performed in the absence or presence of the proteasome inhibitor MG132 (Calbiochem) at 50 M and analyzed by SDS-PAGE as described above. These experiments showed that inhibition of the proteasome resulted in a dramatic
15 increase in the amount of fusion protein synthesized for -lactamase fusions containing 2, 3 or 4 copies of ubiquitinG76V (SEQ. ID. NO. 17) while MG132 had very little or no significant effect on the synthesis of wild-type -lactamase or 1XUb-Bla (FIG. 2B). Use of MG132 in these *in vitro* reactions also revealed the presence of labeled high molecular weight species that represent extended ubiquitin chains conjugated to the ubiquitinG76V- -lactamase fusions (also see
20 Example 16). Therefore, the uncleavable ubiquitinG76V domains (SEQ. ID. NO. 17) in the multiubiquitin destabilization domain may be acting as high affinity conjugation sites for further ubiquitination by E2/E3 ubiquitin ligases. The relative lack of these high molecular weight species in the absence of MG132 reflects the highly efficient recognition and degradation by the proteasome of proteins tagged with extended polyubiquitin chains.

25

30

Example 6. Characterization of multimerized destabilization domain-Naturally Fluorescent Protein fusions *in vitro*.

Characterization of the turnover of multiubiquitin-GFP fusion proteins *in vitro* was
35 similar to the multiubiquitin- -lactamase analyses described in Example 5, except that time points were taken at 0, 30, 60, 90 and 120 min. These experiments showed that Emerald GFP (SEQ. ID. NO. 28) is extremely stable under these conditions, and that the multiubiquitin

destabilization domain was able to impart a short half-life upon the multiUb-GFP fusion proteins (FIG. 3). A striking feature of this analysis was that significant destabilization of GFP required higher numbers of ubiquitinG76V (SEQ. ID. NO. 17) domains than was the case for - lactamase; -lactamase could be strongly destabilized *in vitro* by fusion with as few as two ubiquitinG76V domains (SEQ. ID. NO. 17) (FIG. 2A) whereas GFP required at least three ubiquitinG76V domains (SEQ. ID. NO. 17) to be strongly destabilized (FIG. 3). This relationship between the destabilization domain, and the protein to be destabilized, emphasizes the utility of the multiubiquitin destabilization system, in that the extent of destabilization can be manipulated to give the desired properties by altering the number of ubiquitinG76V (SEQ. ID. NO. 17) domains that are present in the destabilization domain.

Example 7. Characterization of multimerized destabilization domain-endogenous mammalian protein fusions *in vitro*.

Characterization of the turnover of multiubiquitin-caspase-3 fusion proteins *in vitro* was performed as described in Example 5. The TNT synthesis reactions were diluted into chase lysate in the presence of cycloheximide and chase time points were taken and analyzed by SDS-PAGE and autoradiography. FIG. 4 shows that wild-type caspase-3 is stable over a 60 minute chase *in vitro*, and that fusion to the multiubiquitin destabilization domain results in rapid degradation. In particular, the ubiquitinG76V-caspase-3 fusions are degraded in a very similar manner to the ubiquitinG76V- -lactamase fusions although the Ub-caspase-3 fusions appear to be degraded slightly slower *in vitro* than the Ub- -lactamase fusions. Altogether, these data demonstrate the generalized applicability of the multiubiquitin destabilization domain approach to provide for predictable destabilized of any given chosen target protein using this system.

Example 8. Characterization of the half-life of multimerized destabilization domain- - lactamase fusion proteins within cells.

UbiquitinG76V- -lactamase constructs in pcDNA3 (SEQ. ID. NOs. 23 to 26) were introduced into Jurkat T-lymphocytes by electroporation. Stable transfectants were selected in RPMI 1640 media containing 10% fetal bovine serum (Gibco) and 0.8 mg/ml G418 (Geneticin, Gibco). Analysis of -lactamase activity in intact Jurkat cells stably transfected with the pcDNA3-1-4XUb-Bla (SEQ. ID. NOs. 23 to 26) constructs was accomplished by loading the cells with the fluorescent -lactamase substrate CCF2/AM as described in Zlokarnik *et al.* (1998) (Science 279, 1848) followed by analysis by fluorescence activated cell sorter (Becton Dickinson FACS™ Vantage™) or CytoFluor microtiter plate fluorimeter (Perseptive Biosystems). For kinetic measurements, to determine the half-life of the fusion protein *in vivo*,

direct measurements were made of -lactamase activity in lysates prepared from cells expressing the various ubiquitinG76V-Bla fusions.

        Flow cytometry and cell sorting were conducted using a Becton Dickinson FACS™ Vantage™ with a Coherent Enterprise II™ argon laser producing 60mW of 351-364 nm multi-
5   line UV excitation. The flow cytometer was equipped with pulse processing and the Macrosort™ flow cell. Cells were loaded with 1 M CCF2/AM for 1-2 hours at room temperature prior to sorting, and fluorescence emission was detected via 460/50nm (blue) and 535/40nm (green) emission filters, separated by a 490nm long-pass dichroic mirror. The results from one such experiment are shown in **FIG. 5**, where the abundance of cells expressing relatively high levels
10   of -lactamase (regions R5+R6+R7) was determined. This analysis showed that the relative abundance of cells expressing high steady state levels of -lactamase was inversely proportional to the number of copies of ubiquitinG76V (SEQ. ID. NO. 17) fused to -lactamase, i.e., the lowest levels of -lactamase expression were found in cells expressing -lactamase fusions containing the most copies of ubiquitinG76V (SEQ. ID. NO. 17).

15   Similar cytometric analysis experiments were used to investigate the degradation properties of multiUb-Bla fusions *in vivo*. Jurkat cells expressing multiUb-Bla fusions were treated with 50 M MG132 to investigate whether the low -lactamase activity found in cells expressing 3-4XUb-Bla requires proteasome activity. The results, shown in **Table 5**, below show that the addition of inhibitor (+inh/-chx samples) results in a significant increase in the
20   percentage of positive BLA expressing cells for the 2X, 3X and 4X ubiquitinG76V fusion protein constructs compared to the untreated controls (–inh/-chx samples.)

TABLE 5				
	-inh/-chx %Bla$^+$ cells	+inh/-chx %Bla$^+$ cells	-inh/+chx %Bla$^+$ cells	+inh/+chx %Bla$^+$ cells
WT Bla	22.5	22.7	17.6	19.0
1XUb-Bla	17.4	18.8	8.5	16.2
2XUb-Bla	12.0	17.1	2.1	12.2
3XUb-Bla	8.3	14.6	1.5	9.8
4XUb-Bla	4.1	12.1	0.5	5.0

25   Furthermore, treating these cells with 100 g/ml cycloheximide (to block protein synthesis) for one hour prior to CCF2 loading and cytometric analysis (compare columns [–inh/+chx] and [–inh/-chx]) resulted in a strong decrease in -lactamase activity only in cells expressing 2-4XUb-Bla and this decrease could largely be blocked by preincubating the cells with 50 M MG132 prior to cycloheximide addition (column +inh/+chx, in **Table 5**).

These data are strong evidence that the multiubiquitin domain in ubiquitinG76V-Bla fusions is acting as a destabilization motif that directs the rapid degradation of the fusions in a proteasome-dependent manner that is controlled by the number of ubiquitinG76V (SEQ, ID. NO. 17) domains within the multiubiquitin destabilization domain.

5   In order to determine accurate quantitative measurements of the kinetic characteristics of the degradation of UbiquitinG76V- -lactamase fusions *in vivo*, -lactamase activity was determined in cellular lysates. To do this, Jurkat cells expressing the various forms of multiUb-Bla fusion proteins were sorted by flow cytometry to obtain a pool of cells representative of the Bla+ population seen in **FIG. 5** (Region R5+R6+R7). These cells were treated with 100 g/ml
10   cycloheximide to inhibit new protein synthesis, and aliquots of cells were taken at appropriate intervals, to measure the -lactamase activity remaining. This approach enabled a determination of the rate of destruction of the cellular pool of -lactamase fusion proteins within the cell. - lactamase activity was determined in these cell samples by transferring them to ice to terminate further metabolism, and then pelleted by centrifugation. The cell pellets were converted to
15   lysates and -lactamase activity was measured *in vitro* using the free acid form of the - lactamase substrate CCF2. Aliquots of the lysates were assayed using 10 M CCF2 in PBS at room temperature. Hydrolysis of the fluorescent substrate was monitored in a Perseptive Biosystems CytoFluor plate reader using a 395/25nm excitation filter and 460/40nm emission filter.

20   In agreement with the cell analyses by flow cytometry, cells expressing wild-type - lactamase had high levels of -lactamase activity, that was relatively resistant to proteolytic degradation over a 90 minute incubation period with cycloheximide; wild-type -lactamase activity decayed with a half-life >2 hours (**FIG. 6**). Cells expressing 1XUb-Bla fusions also contained relatively high levels of -lactamase activity that decayed with a half-life of about 20-
25   30 minutes. Cells expressing -lactamase fused to 2 or more copies of ubiquitinG76V (SEQ. ID. NO. 17) had significantly less -lactamase activity at steady state (compare 0 minute time points) and the half-lives of these pools of fusion proteins were strikingly short, with all three fusion proteins decaying with *in vivo* half-lives of less than 10 minutes.

The -lactamase measurements from the Jurkat cell lysates allows a calculation of the
30   intracellular concentration and copy number of -lactamase fusion proteins in the respective cell lines. A standard curve created of the hydrolysis of CCF2 by purified -lactamase enzyme was generated and used to calculate the steady state concentration of -lactamase fusion protein for each cell line. This analysis showed that there was a ten-fold difference in intracellular concentration between wild-type -lactamase and 4XUb- -lactamase at steady state (**Table 6**).
35   The calculated concentration of wild-type -lactamase corresponds to 21,000 molecules per cell,

in very good agreement with the values reported by Zlokarnik *et al.* (1998) (Science 279, 1848) for cells expressing high levels of wild-type -lactamase.

TABLE 6		
Construct	Half-life	Intracellular Concentration
WT Bla	>120 min	35 nM
1XUb-Bla	20-30 min	30 nM
2XUb-Bla	<10 min	7 nM
3XUb-Bla	<10 min	5 nM
4XUb-Bla	<10 min	3.5 nM

5   The kinetic data on fusion protein turnover, together with the steady state concentration measurements, demonstrate that the fusion of a multiubiquitin destabilization domain to a target protein allows for the manipulation of both the intracellular concentration, as well as, the turnover kinetics of the resulting fusion proteins. The present invention provides for a method of regulating the intracellular concentration of any target protein within a cell, independently of the
10   rate of transcription of that protein. Unlike other systems of regulating the intracellular concentrations of target proteins, the present invention provides for the ability to "preset" the final concentration of the target protein within a ten-fold range of expression.

The data with multiubiquitinG76V- -lactamase fusions demonstrate that fusions containing one to four copies of ubiquitinG76V fused to -lactamase results in chimeric proteins
15   with half-lives *in vivo* of from 5 to 30 minutes. There are likely to be applications that require proteins that have a half-life longer than that obtained with fusion to one copy of ubiquitinG76V. For such instances, it would be useful to have a form of uncleavable ubiquitin that is recognized by E2/E3 ubiquitin ligases with lower affinity and therefore result in less destabilization than with fusions to ubiquitinG76V. The efficient recognition and degradation of proteins by the
20   proteasome requires the formation of extended polyubiquitin chains that are extended in isopeptide linkage between a critical lysine residue on ubiquitin to the C-terminus of the incoming ubiquitin. The internal lysine in ubiquitin most often used in such polyubiquitin chains is lysine-48. In order to create a longer half-life protein, it is recognized that it is possible to mutagenize the ubiquitin homolog fused to the protein of interest such that it is not recognized
25   by E2/E3 ubiquitin ligases as efficiently as wild-type ubiquitin. It is likely that mutagenesis of lysine-48, (to Arg, His, Gln or Asn for example) and / or the residues surrounding it will yield a form of ubiquitin that is recognized and extended with lower affinity, than the non-mutant forms.

The non extandable homologs would thus serve to create fusion proteins with longer half lives than is otherwise possible with wild-type ubiquitin. Typically such constructs would contain between one and five copies of the non-extendable, non-cleavable ubiquitin homologs to provide for a wide range of destabilization.

5   Alternatively, random mutagenesis of the ubiquitin or mutation of other lysines in ubiquitin may result in a form of ubiquitin with the desired properties.

Example 9. Characterization of the stability of multimerized destabilization domain-Naturally Fluorescent Protein fusions within cells.

10   UbiquitinG76V-GFP constructs in pcDNA3 (SEQ. ID. NOs. 31 to 34) were introduced into CHO cells by Lipofectamine (Life Technologies) transfection. Stable transfectants were selected in RPMI 1640 media containing 10% fetal bovine serum (Gibco) and 0.8 mg/ml G418 (Geneticin, Gibco). Analysis of GFP fluorescence in CHO cells stably transfected with various ubiquitinG76V-GFP constructs was analyzed by flow cytometry on a Becton Dickinson FACS™

15   Vantage™ with a Coherent Enterprise II™ argon laser producing 60mW of 488nm UV excitation. The flow cytometer was equipped with pulse processing and the Macrosort™ flow cell. Fluorescence emission was detected via 530/30nm emission filter. The FACS analyses of stable populations determined that the steady state percentage of bright green GFP^+ cells varied depending on the presence of the multiubiquitin destabilization domain. The relative percentages

20   of GFP^+ cells are shown in the **Table 7**.

TABLE 7	
Stable CHO cell line	**% GFP^+ cells**
Wild-type GFP	39.13
1XUb-GFP	5.74
2XUb-GFP	3.06
3XUb-GFP	2.2
4XUb-GFP	1.93

This analysis showed that the relative abundance of cells expressing high steady state levels of GFP fluorescence was inversely proportional to the number of copies of ubiquitinG76V (SEQ. ID. NO. 17) fused to the protein, i.e., the lowest levels of GFP-expressing cells were

25   found in the fusions containing the most copies of ubiquitinG76V (SEQ. ID. NO. 17). The steady state concentration measurements demonstrate that fusions of a multiubiquitin destabilization domain to the highly stable GFP mutant Emerald (SEQ. ID. NO. 28) allows for the predictable

and controllable manipulation of the intracellular concentrations of naturally fluorescent proteins.

Example 10. Construction of destabilization domain – linker – reporter moiety fusion proteins

Ubiquitin- -lactamase fusion proteins containing a specific protease cleavage site were constructed by annealing the complementary oligonucleotides DEVD-1 (SEQ. ID. NO. 45) and DEVD-2 (SEQ. ID. NO. 46) that encode a caspase-3-type cleavage site and produce BamH I compatible ends. This oligonucleotide cassette was ligated into BamH I-digested pcDNA3-1-4XUb-Bla plasmid constructs (SEQ. ID. NOs. 23 to 26) described in Example 2. The resulting constructs encode an in-frame fusion protein consisting of from one, to four, copies of ubiquitinG76V (SEQ. ID. NO. 17) separated from -lactamase by linker containing a caspase-3 cleavage site; the plasmids were designated as pcDNA3-1-4XUb-DEVD-Bla (SEQ. ID. NOs. 47-50). A control linker containing a DEVA site that should not serve as a cleavage site for caspase-3-like proteases was constructed in an identical manner using DEVA1 (SEQ. ID. NO. 51) and DEVA2 primers (SEQ. ID. NO. 52) and the resulting plasmids were designated as pcDNA3-1-4XUb-DEVA-Bla (SEQ. ID. Nos. 53-56).

End of Ubiquitin-G76V (SEQ. ID. NO. 17)   Start of -lactamase

LVLRLRGVGSVGAVGSVGDEVDGSGAWLHPETLVKV

Recognition site for post-translational activity

Example 11. Detection of caspase activity using destabilized reporter moieties *in vitro*

[35]S-labeled ubiquitin- -lactamase fusion proteins containing a cleavage site for the group II effector caspase-3 were produced by *in vitro* transcription/translation reactions as described in Example 2 except that plasmids pcDNA3-1-4XUb-DEVD-Bla (SEQ. ID. NOs. 47-50) or control plasmids pcDNA3-1-4XUb-DEVA-Bla (SEQ. ID. NOs. 53-56) were used as templates. The ^{35}S-labeled proteins were then used as substrates for purified caspase-3 in an *in vitro* cleavage reaction. The 12 l reaction consisted of 4 l of ^{35}S-labeled ubiquitin-DEVD/Æla fusion proteins, 100 mM HEPES pH 7.5, 10% sucrose, 0.1% CHAPS, 10 mM DTT and 25 nM purified recombinant caspase-3. The reactions were incubated at 30°C and samples taken at 0, 5, 10, 20,

30, 45, and 60 minutes and analyzed by SDS-PAGE and autoradiography. The results from 2XUb-DEVD-Bla and 2XUb-DEVA-Bla fusion proteins are shown in FIG. 7A. The 2XUb-DEVD-Bla fusion served as a very good substrate for caspase-3 with over 90% cleavage within 5 minutes. In contrast, the 2XUb-DEVA-Bla fusion was not cleaved by caspase-3 *in vitro*, even at extended incubation times. The 2XUb-DEVD-Bla cleavage product seen in FIG. 7A co-migrates on SDS-PAGE gels with -lactamase fused to the short DEVD linker region (data not shown) and verifies the position of the cleavage site and identifies the labeled cleavage product as the -lactamase portion of the cleaved fusion. The liberated destabilization domain is much smaller and has run off the gel in this experiment. These data demonstrate that the DEVD fusion serves as an efficient substrate for caspase-3 and the lack of cleavage with the DEVA fusion confirms that the cleavage is occurring ate the DEVD site.

The protease assay outlined above requires that the protease cleavage result in a stabilization of the catalytic domain of the reporter. To test whether this is the case, we mixed approximately equal portions of cleaved and uncleaved ^{35}S-labeled reporters from *in vitro* cleavage reactions identical to those in FIG. 7A and then diluted the fragments into crude chase lysate containing cycloheximide to perform a chase experiment. The reactions were incubated at 37°C and samples were taken at 0, 5, 10, 20, 30 and 60 minutes and analyzed by SDS-PAGE and autoradiography. FIG. 7B shows that the uncleaved intact 2XUb-DEVD-Bla or 2XUb-DEVA-Bla reporters were degraded very rapidly *in vitro* with a half-life of less than 5 minutes. In contrast, the cleavage product from the 2XUb-DEVD-Bla reporter lacks the destabilization domain and as a result is very stable *in vitro*. These data confirm that the intact and cleaved versions of the -lactamase reporters have dramatically different half-lives and provide evidence that this difference in stability may provide a format for assaying endoprotease activity *in vivo*.

Example 12. Detection of effector caspase protease activity using destabilized reporter moieties within cells

Plasmids pcDNA3-1-4XUb-DEVD-Bla (SEQ. ID. NOs. 47-50) and pcDNA3-1-4XUb-DEVA-Bla (SEQ. ID. NOs. 53-56) were transfected into Jurkat cells and selected for stable transfectants as described in Example 8. The stable transfectants were sorted by flow cytometry using Becton Dickinson FACS™ Vantage™ SE and FACS™ Vantage™ flow cytometers. The FACS™ Vantage™ SE was equipped with Turbosort Option, pulse processing, ACDU, and Coherent Innova 302C krypton and Coherent Innova 70 Spectrum mixed-gas krypton-argon lasers. The FACS™ Vantage™ was equipped with pulse processing, ACDU, and Coherent Enterprise II and Coherent Innova 70 Spectrum mixed-gas krypton-argon (with violet option) lasers. For β-lactamase experiments, 60mW of 413nm laser emission was used for CCF2

64

excitation, with a 500nm dichroic filter separating a 460/50nm (CCF2 blue fluorescence) and a 535/40nm bandpass filter (green fluorescence). Single cells with the desired level of -lactamase expression were sorted into individual wells of 96-well plates using the Automatic Cell Deposition Unit (ACDU) on the FACS™ Vantage™ and expanded for analysis of

5   homogeneous clonal populations. All results in this Example utilized clonal lines.

The clonal cell lines were initially screened for expression of -lactamase and the ability to degrade the Ub-DEVD-Bla or Ub-DEVA-Bla fusion rapidly. This initial screen was accomplished by treating an aliquot of cells with 100 g/ml cycloheximide followed by incubation at 37°C for 1 hour (chase period). Treated and untreated cells were loaded with 1

10   M CCF2-AM for 1 hour at room temperature and -lactamase levels were quantified using a CytoFluor microtiter plate fluorimeter (Perseptive Biosystems) using 395/25nm excitation and 460/40 (blue) nm and 530/30 (green) nm emission filters. Emission ratios were calculated from background-subtracted values (background=media+CCF2 alone) and expressed as a 460/530nm ratio where a high ratio indicates high -lactamase activity. This analysis showed that Ub-

15   DEVD-Bla fusions with two or more copies of ubiquitinG76V (SEQ. ID. NO. 17) gave satisfactory chase characteristics, with fusions to two copies of UbiquitinG76V (SEQ. ID. NO. 17) giving the highest steady state levels (no chase) of fusion protein (data not shown). In contrast, 1XUb-DEVD-Bla fusions were not sufficiently destabilized to be usable with this assay format as cells expressing the fusion required extended cycloheximide treatments (data not

20   shown). As the 2-4XUb-DEVD-Bla fusions all exhibited satisfactory rates of proteolytic turnover in cells, the 2X ubiquitinG76V destabilization domain was used with the DEVD-Bla fusions because it gave the best performance (expression levels vs. turnover kinetics) in this particular application. It is worth noting here that due to the variability in the intrinsic stability of different proteins fused to the ubiquitinG76V (SEQ. ID. NO. 17) destabilization domain;

25   fusions of other cellular proteins with multimerized destabilization constructs would be expected to require a dissimilar number of copies of ubiquitinG76V (SEQ. ID. NO. 17) to impart sufficiently rapid turnover kinetics (data not shown). A key advantage of the present invention is the ability to meet this need by varying the number of destabilization domains present within the multimerized destabilization domain construct.

30   One clonal cell line from each of 2XUb-DEVD-Bla and 2XUb-DEVA-Bla cell populations was characterized in detail. To establish the background (no -lactamase) control value, wild-type Jurkat cells containing no -lactamase activity were loaded with CCF2-AM and the 460/530 fluorescence ratio measured. The value obtained, about 0.05, establishes the background ratio exhibited by cells in the absence of -lactamase activity. When the 2XUb-

35   DEVD-Bla and 2XUb-DEVA-Bla clones were treated with cycloheximide (chx) for 1 hour at

37°C prior to CCF2-AM loading, they both exhibited 460/530 ratios very near the background ratio of 0.05, demonstrating that the cells retained the ability to degrade the 2XUb-Bla fusion very efficiently **(Table 8)**.

TABLE 8		
	2XUb-DEVD-Bla 460/530 emission ratio	**2XUb-DEVA-Bla** 460/530 emission ratio
no chx	1.80	1.60
+ chx	0.07	0.07
+ Fas/-chx	1.25	1.10
+ Fas/+chx	0.67	0.12
+ Fas/+inh/+chx	0.08	0.09

The fact that there is a significant difference in stability between the uncleaved reporter and the cleavage product *in vitro* **(FIG. 7B)** forms the basis for an assay for protease activity in intact cells. As shown in **Table 8**, in the absence of caspase activity, both 2XUb-DEVD-Bla and 2X-Ub-DEVA-Bla fusions are rapidly degraded to very low levels in the presence of cycloheximide to inhibit new protein synthesis. Treatment of Jurkat cells with Fas ligand will result in the activation of Fas receptor - an apoptosis signaling receptor found on the surface of a number of cell types that belongs to the tumor necrosis factor (TNF)/nerve growth factor family. Fas activation ultimately leads to the activation of the group II caspases that efficiently cleave substrates containing DEVD recognition motifs. In order to activate this pathway and measure the activity of group II caspases using the DEVD-Bla reporter in intact cells, an anti-Fas antibody (CH-11 anti-Fas IgM; Kamiya Biomedical Co., Seattle, WA) was used to cross-link the receptor and stimulate the activation of group II caspases. Western blot analysis of the anti-Fas-treated cells confirmed the proteolytic activation of caspase-3 (data not shown), the major group II caspase activity in Jurkat cells. Treatment of Jurkat cells expressing 2XUb-DEVD-Bla or 2XUb-DEVA-Bla reporter with 50 ng/ml anti-Fas IgM for 6 hours at 37°C resulted in a modest decrease in the steady-state levels of the reporter **(Table 8)**, most likely due to the inhibition of protein synthesis that is known to accompany apoptosis. At this point, the activation of group II caspases will result in the cleavage and stabilization of some proportion of the DEVD-Bla (but not the control DEVA-Bla) reporters. Treatment of such cells with cycloheximide would then allow for the clearing of the uncleaved, short half-life reporters while leaving the stable cleaved reporters as the sole forms of -lactamase activity in the cells. **Table 8** shows that cycloheximide addition to anti-Fas treated cells (+ Fas/+chx) resulted in the stabilization of a significant fraction of the DEVD-Bla reporters while the DEVA-Bla reporters cannot be cleaved

and stabilized. To show that the stabilization of the DEVD-Bla reporters is due to caspase activation, we used the peptide inhibitor Z-VAD-fmk (Enzyme Systems Products, Livermore, CA) that is a potent broad inhibitor of caspases. Treatment of the cells with 10 M Z-VAD-fmk coincident with anti-Fas addition blocked the stabilization of DEVD-Bla reporters. Treatment of

5   the cells with cycloheximide resulted in the degradation of the non cleaved constructs to background levels of -lactamase activity (+ Fas/+Inh/+chx). Comparison of -lactamase levels in antiFas-treated DEVD-Bla-expressing cells in the presence or absence of Z-VAD-fmk inhibitor determines the dynamic range of the assay; in this particular experiment the dynamic range is approximately 8-fold. These data demonstrate that the cleavage and stabilization of

10   short half-life -lactamase protease reporters provides a sensitive and specific assay for measuring the activation of caspases in intact cells.

It is of note that this assay format would permit the identification of compounds that stimulate group II caspases and subsequent apoptosis (agonist/inducer format) as well as compounds that inhibit caspase activity stimulated by a known reagent such as anti-Fas IgM

15   (antagonist/inhibitor format). As evidence for this assay being useful for both inducer and inhibitor applications, we generated dose-response curves for both an inducer of caspases and apoptosis (anti-Fas IgM) and an inhibitor of anti-Fas induced apoptosis (Z-VAD-fmk). **FIG. 8** shows that the assay in Jurkat cells expressing 2XUb-DEVD-Bla generates sufficient dynamic range to detect low concentrations of the inducer anti-Fas IgM (EC_{50}= 11 ng/ml). In addition,

20   treatment of Jurkat cells expressing 2XUb-DEVD-Bla with 50 ng/ml anti-Fas IgM allows sensitive detection of inhibition by Z-VAD-fmk with IC_{50}= 5 M (**FIG. 8**).

Example 13. Creation of reporters for viral self-cleaving proteases using multimerized

25   **destabilization domain- -lactamase-rhinovirus 2A protease fusions.**

The gene encoding the human rhinovirus 14 2A protease (SEQ. ID. NO. 57) was isolated by PCR amplification from genomic RNA by RT-PCR using oligonucleotides HRV145' (SEQ. ID. NO. 58) and HRV143', (SEQ. ID. NO. 59). The resulting PCR product had BamH I sites at both ends of the HRV14 2A protease sequence and could be inserted in frame into the pcDNA3-

30   1-4XUb-Bla vectors (SEQ. ID. Nos. 23-26) from example 2. The PCR fragment from this reaction was digested with BamH I and ligated into pcDNA3-3XUb-Bla (SEQ. ID. NO. 25). The resulting construct, pcDNA3-3XUb-Bla HRV14 (SEQ. ID. NO. 60) was further characterized *in vitro* and within cells.

In addition to the HRV14 2A protease constructs, two additional constructs were made

35   for the HRV16 2A protease. The gene for the human rhinovirus 16 sequence 2A protease (SEQ. ID. NO. 61) was isolated by polymerase chain reaction (PCR) amplification of a plasmid

template. The PCR template was a plasmid construct containing the entire HRV16 genome (a gift from Dr. Wai Ming Lee at the University of Wisconsin). Oligonucleotides HRV165' (SEQ. ID. NO. 62) and HRV163', (SEQ. ID. NO. 63) were used in a PCR reaction with the HRV16 plasmid resulting in a PCR product that had BamH I sites at both ends of the HRV16 2A

5   protease sequence. The PCR fragment from this reaction was digested with BamH I and ligated into pcDNA3-3XUb-Bla (SEQ. ID. NO. 25) and pcDNA3-Ub-Met Bla (SEQ. ID. NO. 27) via the BamH I site. The resulting constructs were pcDNA3-3XUb-Bla HRV16 (SEQ. ID. NO. 64) and pcDNA3-Ub-Met-Bla HRV16 (SEQ. ID. NO. 65). In addition, two mutant constructs were made for the HRV16 2A protease. These mutants corresponded to mutations at two residues of

10   the putative catalytic triad for the 2A protease and should result in a catalytically inactive mutant, specifically, aspartate 35 was mutated to alanine (D35A) and cysteine 106 was mutated to alanine (C106A). These derivatives were generated by mutagenesis of the HRV16 2A protease using oligonucleotide HRV16 D35A (SEQ. ID. NO. 66) and oligonucleotide HRV16 C106A (SEQ. ID. NO. 67). The resulting plasmids were designated as pcDNA3-3XUb-Bla

15   HRV16(C106A) (SEQ. ID. NO. 68), pcDNA3-3XUb-Bla HRV16(D35A) (SEQ. ID. NO. 69), pcDNA3-Ub-Met-Bla HRV16(C106A) (SEQ. ID. NO. 70) and pcDNA3-Ub-Met-Bla HRV16(D35A) (SEQ. ID. NO. 71).

20   **Example 14. Detection of Rhinovirus protease activity using destabilized reporter moieties in vitro.**

^{35}S-labeled ubiquitin- -lactamase fusion proteins containing the HRV14 and HRV16 2A proteases, as well as the mutants above, were produced by *in vitro* transcription/translation reactions as described in Example 5. The plasmids pcDNA3-3XUb-Bla HRV16 (SEQ. ID. NO.

25   64), pcDNA3-3XUb-Bla HRV16(C106A) (SEQ. ID. NO. 68), pcDNA3-3XUb-Bla HRV16(D35A) (SEQ. ID. NO. 69), pcDNA3-Met Ub-Bla HRV16 (SEQ. ID. NO. 65), pcDNA3-3XUb-Bla HRV14 (SEQ. ID. NO. 60), and pcDNA3-MetUb-Bla HRV14 (SEQ. ID. NO. 72) were used as templates. The reactions were incubated at 30°C for 45 min and analyzed by SDS-PAGE and autoradiography. **FIG. 9A** shows the results of TNT synthesis reactions for the wild-

30   type HRV16 2A and the two mutant HRV16 2A constructs. Shown are the levels of expression for the stable (Met) and destabilized 3X ubiquitinG76V HRV16 2A-Bla fusions. As expected, the level of expression is higher in the stable methionine containing constructs than the destabilized 3XUb constructs (**FIG. 9A**). The wild-type HRV16 2A fusions also show significant accumulation of the lower molecular weight stable cleavage product indicating that

35   the fusions exhibit robust autocatalytic cleavage activity in these *in vitro* reactions. In contrast, mutation of residues in the putative catalytic triad (aspartate 35 and cysteine 106) blocked

formation of the stable cleavage product, indicating that these mutants are indeed catalytically inactive.

The protease assay outlined in Example 10 requires that protease cleavage results in a stabilization of the catalytic domain of the reporter. To test for this requirement the pcDNA3-
5   3XUb-Bla HRV14 TNT reaction was diluted into chase lysate containing cycloheximide to perform a chase experiment. The reactions were incubated at 37°C for 60 minutes and analyzed by SDS-PAGE and autoradiography. **FIG. 9B** shows that the uncleaved intact 3XUb-HRV14-Bla reporter was completely degraded during the 60 minute chase. In contrast, the cleavage product from the 3XUb-HRV14-Bla reporter lacks the destabilization domain, and as a result, is
10   stable *in vitro*. These data confirm that the intact and cleaved versions of the HRV 2A--lactamase fusion reporters have dramatically different half-lives and provide evidence that this difference in stability can provide the basis for assaying self-cleaving protease activity in side intact cells.

15   **Example 15.   Detection of Rhinovirus protease activity using destabilized reporter moieties**
in vivo.

The biochemical properties of self-cleaving *cis* proteases such as rhinovirus 2A pose several technical challenges that have hampered the development of a screening format to allow for the identification of inhibitors or activators in cell based assays. First, the activity of the
20   protease is directed toward cleavage of itself. This rules out the use of separate reporters that are cleaved in *trans* and limit the catalytic output of the assay, i.e., a single protease molecule generates a single cleavage product and this fact eliminates the catalytic amplification used in traditional assays for trans-cleaving proteases. In order to address these limitations, the -lactamase reporters are incorporated into the 2A protease itself, thereby measuring the *cis*
25   cleavage reaction directly and gaining the advantage of a catalytic reporter that can cleave many CCF2 substrate molecules per reporter. Since the HRV 2A protease undergoes the self-cleavage reaction immediately upon synthesis, the screening assay must be performed on newly synthesized HRV 2A- -lactamase reporters. A screen to identify inhibitors of the protease must incorporate a step where test compounds are added and their effect then measured. As cleaved
30   stable -lactamase reporters will accumulate in the cell as the HRV 2A-Bla reporters are being constitutively expressed, it is essential to eliminate the readout due to such cleavage products that are generated before the test compound is added. To do this, the -lactamase inhibitor clavulanate was used. Clavulanate is a non-cytotoxic irreversible inhibitor of -lactamase and overnight treatment of Jurkat cells reduces -lactamase levels to background (See commonly
35   owned U.S. Patent Application No. 09/067,612 filed April 28 1998). Therefore, clavulanate

treatment of Jurkat cells expressing HRV 2A-Bla fusions eliminates the -lactamase activity that is present in the cell resulting from both uncleaved and cleaved -lactamase reporters. In essence, this has the effect of "zeroing out" the -lactamase activity in the cells and bringing the cells back down to baseline activity. The clavulanate can then be washed out and test compound

5   added. New synthesis of HRV 2A-Bla reporters will result in the accumulation of the fusion protein reporter in the cells and the self-cleavage reaction will now be subject to inhibition by the test compound. After an appropriate interval to allow for the cleavage of newly synthesized reporters has passed, the cells can be treated with cycloheximide to clear out the unstable uncleaved reporters and the resulting -lactamase activity will be due exclusively to cleaved,

10   stabilized reporters.

Plasmids pcDNA3-3XUb-Bla HRV16 (SEQ. ID. NO. 64) and pcDNA3-3XUb-Bla HRV14 (SEQ. ID. NO. 60) were transfected into Jurkat cells and selected for stable transfectants as described in Example 8. The stable transfectants were sorted by flow cytometry using Becton Dickinson FACS™ Vantage™ SE and FACS™ Vantage™ flow cytometers. The FACS™

15   Vantage™ SE was equipped with Turbosort Option, pulse processing, and Coherent Innova 302C krypton and Coherent Innova 70 Spectrum mixed-gas krypton-argon lasers. The FACS™ Vantage™ was equipped with pulse processing, and Coherent Enterprise II and Coherent Innova 70 Spectrum mixed-gas krypton-argon (with violet option) lasers. For β-lactamase experiments, 60mW of 413nm laser emission was used for CCF2 excitation, with a 500nm dichroic filter

20   separating a 460/50nm (CCF2 blue fluorescence) and a 535/40nm bandpass filter (CCF2 green fluorescence). Single cells with the desired level of -lactamase expression were sorted into individual wells of 96-well plates using the Automatic Cell Deposition Unit (ACDU) on the FACS™ Vantage™ and expanded for analysis as homogeneous clonal populations. All results in Example 15 utilized clonal lines.

25   Selected clones (25-50 for each construct) were then expanded further for analysis. Clones were treated for 16 hours with 300 M clavulanate, washed twice with phosphate buffered saline (PBS), incubated for 2 hours at 37°C, treated for 1 hour at 37°C with 100 g/ml cycloheximide, and then loaded with CCF2-AM for 2 hours at room temperature. The individual clones were then screened visually by fluorescence microscopy. At least 24 individual clones

30   were tested in this manner for each construct and one clone chosen for each construct.

To assay HRV 2A protease activity, the selected Jurkat stable cell clones were treated for 16 hours with 300 M clavulanate to inactivate pre-existing cleaved and uncleaved HRV 2A-Bla fusion protein. Cells were then washed twice with PBS, resuspended at 100,000 cells/well in 100 l RPMI + 10% FBS in 96-well plates. The cells were incubated at 37°C for 4 hours in the

35   presence or absence of an inhibitor of the 2A protease. Cells were treated with 100 g/ml

cyclohexamide for 30 minutes at 37°C, loaded with CCF2-AM for 2 hours at room temperature and read on the CytoFluor plate reader as described in Example 8. Inhibitor compounds, radicicol and geldanamycin, were used for the validation of the HRV protease cell-based assay. These compounds are known inhibitors of the Hsp90 heat shock protein (see Roe et al., (1999) J.

5   Med. Chem. 42 260-266), which is required for the folding and regulation of a number of cellular proteins and can inhibit HRV 2A protease activity *in vitro* (data not shown). Compounds were tested at 1 M for their ability to inhibit the HRV 2A protease cell-based assay using clones expressing HRV16 and HRV14 2A protease reporters. Jurkat cells expressing 3XUb-Bla-HRV14 or HRV16 2A protease fusion proteins contained significant -

10   lactamase activity in the absence of the inhibitors (Table 9). Both radicicol and geldanamycin showed strong inhibition of cellular -lactamase activity remaining after the cyclohexamide chase. The inhibitors are not simply inhibiting -lactamase enzyme activity because control experiments showed that radicicol and geldanamycin did not inhibit -lactamase activity in Jurkat cells expressing wild-type -lactamase (data not shown). These data demonstrate that the

15   -lactamase activity present after a cyclohexamide chase is due to HRV 2A protease activity and that this -lactamase activity can be blocked using inhibitors of HRV 2A protease. These results further demonstrate that Jurkat cells expressing 3XUb-Bla HRV 2A fusion proteins constitute a robust cell-based assay for HRV 2A cis-protease activity. The difference in -lactamase activity between untreated and inhibitor-treated cells determines the dynamic range of

20   this assay; in this particular experiment, the assay dynamic range is approximately 6-fold.

TABLE 9		
	3XUb-HRV14-Bla 460/530nm ratio	3XUb-HRV16-Bla 460/530nm ratio
no inhibitor	1.022	0.895
+ radicicol	0.152	0.229
+ geldanamycin	0.153	0.239

Example 16. Detection of Proteasome activity within cells using destabilized reporter moieties and use in the identification of proteasome inhibitors.

25   A direct application of the destabilized reporter fusions is in the measurement of the activity of the proteolytic activity that responsible for the constitutive degradation of the reporter in cells. Ubiquitinated proteins are known to be degraded by the multi-subunit proteasome. In addition, the proteasome is responsible for the degradation of the large majority of cellular proteins see Lee and Goldberg, (1998) Trends Cell Biol., 8 397-403). The proteasome itself has

30   been implicated in a number of pathological conditions resulting from either increased or

decreased proteasome activity (see Ciechanover, (1998) EMBO J. 17 7151-7160). As such, the proteasome represents an attractive target for intervention in pathological conditions using small molecule inhibitors or activators.

Inhibitors of the proteasome were initially tested *in vitro* for inhibition of degradation of
5   2XUb-Bla. Transcription/translation reactions on the pcDNA3-2XUb-Bla (SEQ. ID. NO. 24) construct were preformed as described in Example 5. The ^{35}S-labeled synthesis reactions were diluted into crude chase lysates in the presence of cycloheximide and inhibitor and incubated at 37°C for 20 minutes. Samples were then analyzed by SDS-PAGE and autoradiography. FIG. 10 shows that >90% of the starting ^{35}S-labeled fusion protein is degraded by the 20 minute time
10   point in the absence of proteasome inhibitors. Addition of the inhibitor MG132 (Calbiochem) at 50 M resulted in a significant increase in the intact, un-conjugated fusion protein as well as the appearance of high molecular weight labeled species that represent extensive further ubiquitination of the fusion protein. The high molecular weight ubiquitin conjugates accumulate prominently in the presence of MG132 because they are recognized so efficiently by the
15   proteasome that they are barely visible without inhibiting their degradation.   Additional proteasome inhibitors gave very similar results: 10 M lactacystin -lactone (Calbiochem) and 50 M Ac-LLN (Sigma) stabilized the 2XUb-Bla fusion protein and caused the accumulation of high molecular weight ubiquitin conjugates.

Proteins destined to be degraded by the proteasome are initially modified by the covalent
20   addition of ubiquitin to lysines within the targeted protein through an isopeptide linkage between the C-terminal residue of ubiquitin and the -amino groups of the substrate protein.   The conjugated ubiquitin(s) acts as a high affinity conjugation site for the addition of additional ubiquitin polypeptides in isopeptide linkage between the C-terminus of the incoming ubiquitin to a lysine residue within the conjugated ubiquitin. When the ubiquitin chains reach a critical size
25   four or more ubiquitin residues long (see Thrower et al., (2000) EMBO J. 19 94-102)), the ubiquitin-protein conjugate is recognized by the proteasome with high affinity, the substrate protein is degraded and the ubiquitin residues are recycled for further rounds of ubiquitination. To test whether poly-ubiquitination is required for the degradation of 2XUb-Bla, we used a form of ubiquitin where all amines had been reductively methylated, thereby producing a form of
30   ubiquitin that can be conjugated but not extended. When methylated ubiquitin (MeUb) was added to the *in vitro* degradation system at 1 mg/ml, it significantly stabilized 2XUb-Bla and resulted in the appearance of ladders of labeled species that contain low (1-5 copies) numbers of conjugated ubiquitin polypeptides. (FIG. 10) It also inhibited the formation of the very high molecular weight ubiquitin-substrate conjugates observed with the proteasome inhibitors.
35   Collectively, the *in vitro* inhibitor data demonstrate that the multiubiquitin destabilization

domain targets degradation of the protein it is fused to in a proteasome-dependent manner that requires poly-ubiquitination of the substrate for high efficiency recognition/degradation.

Jurkat cells expressing 2XUb-Bla fusion protein were used to test several inhibitors of proteasome function that were active in the *in vitro* system to determine if they were also active
5 within living cells. Cells were treated with various concentrations of the proteasome inhibitors MG132 or Ac-LLN for 30 minutes at 37°C and then cycloheximide was added to 100 g/ml to initiate a chase period. After 1 hour at 37°C, the cells were cooled to room temperature and then loaded with 1 M CCF2-AM and -lactamase activity quantified using a CytoFluor plate reader. The background-subtracted emission values at 460 nm and 530 nm were expressed as a
10 460/530 ratio and dose-response curves were plotted. **FIG. 11** shows that both MG132 and Ac-LLN exhibited a dose-dependent inhibition of the decay of -lactamase activity indicating that they had inhibited the intracellular degradation of the ubiquitin- -lactamase fusion protein. IC_{50} values calculated from linear regression analysis were found to be 13 M for Ac-LLN and 2.1 M for MG132 and are within the characteristic range for inhibition of substrates degraded by
15 the proteasome (see Lee and Goldberg, (1998) Trends Cell Biol., 8 397-403). These data demonstrate that the multiubiquitin destabilization domain fused to -lactamase can serve as a robust cell-based 96-well format screening assay for inhibitors of the proteasome.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the
20 invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or
25 related fields are intended to be within the scope of the following claims.

CLAIMS

We claim.

1. A method of detecting an activity in a cell, comprising:

5   1) providing a cell comprising,

           a)   at least one destabilization domain, wherein said destabilization domain is non-cleavable by -NH-ubiquitin protein endoproteases,

           b)   a reporter moiety, and

           c)   a linker moiety that operatively couples said destabilization domain to said reporter moiety,

10   wherein said linker moiety comprises a recognition motif for said activity and modification of said linker moiety by said activity modulates the coupling of said destabilization domain to said reporter moiety thereby modulating the stability of said reporter moiety, and

        wherein said linker moiety is non-cleavable by said -NH-ubiquitin protein endoproteases,

15   2) detecting said reporter moiety, or a product of said reporter moiety.

2. The method of claim 1, wherein said at least one destabilization domain is arranged as linear multimer, and

        wherein said linear multimer comprises at least two copies of said destabilization domain and is non-cleavable by said -NH-ubiquitin protein endoproteases.

20

3. The method of claim 1, wherein said linker moiety is non-naturally occurring polypeptide or protein.

4. The method of claim 1, wherein said linker moiety covalently couples said destabilization domain to said reporter protein.

25

5. The method of claim 1, wherein said linker moiety is between about 1 and 30 amino acid residues.

6. The method of claim 1, wherein said destabilization domain comprises a ubiquitin homolog.

30 7. The method of claim 6, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by said -NH-ubiquitin protein endoproteases.

8. The method of claim 6, wherein said ubiquitin homolog comprises a mutation at glycine 76.

35 9. The method of claim 1, wherein said linker moiety comprises a first amino acid sequence that is covalently coupled to said reporter moiety, and a second amino acid sequence that is covalently coupled to said at least one destabilization domain.

74

10. The method of claim 1, wherein said activity is selected from the group consisting of a protease activity, a protein kinase activity and a phosphoprotein phosphatase activity.

5   11. The method of claim 1, wherein said reporter moiety is selected from the group consisting of a naturally fluorescent protein homolog, a -lactamase homolog, a -galactosidase homolog, an alkaline phosphatase homolog, a CAT homolog, and a luciferase homolog.

12. The method of claim 11, wherein said reporter moiety comprises a -lactamase homolog.

10

13. The method of claim 11, wherein said reporter moiety comprises an *Aequorea* Green fluorescent protein homolog.

14. The method of claim 11, wherein said reporter moiety comprises an Anthozoan Green fluorescent protein homolog.

15 15. The method of claim 1, wherein said cell is a mammalian cell.

16. The method of claim 1, wherein said cell is a yeast cell.

17. The method of claim 1, wherein said cell is an insect cell.

20

18. The method of claim 1, wherein said cell is a plant cell.

19. The method of claim 1, wherein said method further comprises the step of adding a protein synthesis inhibitor to said cell.

25

20. The method of claim 1, wherein said method further comprises the step of adding an inhibitor of said reporter moiety to said cell.

21. The method of claim 1, wherein said method further comprises the step of adding a test chemical to said cell.

30

22. The method of claim 20, wherein said method further comprises the step of relating said reporter moiety activity before addition of said test chemical to said reporter moiety activity after addition of said test chemical.

23. A method of regulating the concentration of one or more target proteins in a cell, comprising;

35   1)   providing a cell comprising,

a) a linear multimerized destabilization domain, wherein said linear multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at least two copies of a destabilization domain,

b) a target protein, and

5

c) a linker that operatively couples said linear multimerized destabilization domain to said target protein,

wherein said linker comprises a protease cleavage site for a protease and cleavage of said linker by said protease modulates the coupling of said linear multimerized destabilization domain to said target protein, thereby modulating the stability of said target protein in said cell,

10

and

wherein said linker is non-cleavable by a -NH-ubiquitin protein endoproteases,

2) providing said protease to cause cleavage of said linker thereby increasing the stability and concentration of said protein of interest in said cell.

15   24. The method of claim 23, wherein said protease is naturally expressed in said cell.

25. The method of claim 23, wherein said protease is not naturally expressed in said cell.

26. The method of claim 23, further comprising the step of adding an inhibitor of said protease.

20

27. The method of claim 23, wherein said linker is between 1 and 30 amino acid residues.

28. The method of claim 23, wherein said cell is a mammalian cell.

25   29. The method of claim 23, wherein said cell is a yeast cell.

30. The method of claim 23, wherein said cell is an insect cell.

31. The method of claim 23, wherein said destabilization domain comprises a ubiquitin homolog.

30

32. The method of claim 31, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases.

33. The method of claim 31, wherein said ubiquitin homolog comprises a mutation at glycine 76.

35

34. The method of claim 23, wherein said protease is provided by transfecting said cell with an expression vector comprising a nucleic acid sequence encoding said protease.

76

35. The method of claim 34, wherein said expression vector further comprises an inducible promoter.

36. The method of claim 34, wherein said expression vector is a retroviral expression vector.

5

37. The method of claim 34, wherein said protease is a viral protease.

38. A method of destabilizing a target protein in a cell, comprising;

     operatively coupling a target protein to a linear multimerized destabilization domain, wherein said linear

10   multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at

     least two copies of a destabilization domain.

39. The method of claim 38, wherein said destabilization domain comprises a ubiquitin homolog.

15   40. The method of claim 39, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin

     protein endoproteases.

41. The method of claim 39, wherein said ubiquitin homolog comprises a mutation at glycine 76.

20   42. The method of claim 38, wherein said protein of interest is fused in frame to said multimerized destabilization domain.

43. The method of claim 38, wherein said protein of interest is non-covalently coupled to said multimerized ubiquitin fusion

     protein.

25   44. The method of claim 38, wherein said cell is a mammalian cell.

45. The method of claim 38, wherein said cell is a yeast cell.

46. The method of claim 38, wherein said cell is an insect cell.

30

47. The method of claim 38, wherein said cell is a plant cell.

48. The method of claim 38, wherein said target protein is coupled to said multimerized destabilization domain by a linker.

35   49. The method of claim 48, wherein said linker is between 1 and 10 amino acid residues.

50. A recombinant DNA molecule, comprising a nucleic acid sequence encoding for;

     a)   a linear multimerized destabilization domain, wherein said linear multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at least two copies of a destabilization domain,

     b)   a target protein, and

5   c)   a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

            wherein said linker is non-cleavable by a -NH-ubiquitin protein endoproteases.

51. The method of claim 50, wherein said linker moiety comprises an enzyme modification site for an activity, and
10   modification of said linker moiety by said activity modulates the coupling of said multimerized destabilization domain to said reporter moiety, thereby modulating the stability of said reporter moiety.

52. The method of claim 50, wherein said destabilization domain comprises a ubiquitin homolog.

15   53. The method of claim 52, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases.

54. The method of claim 52, wherein said ubiquitin homolog comprises a mutation at glycine 76.

20   55. A recombinant protein molecule, comprising an amino acid sequence encoding for;

     a)   a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

     b)   a target protein, and

25   c)   a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

            wherein said linker is non-cleavable by a -NH-ubiquitin protein endoproteases.

56. The method of claim 55, wherein said linker moiety comprises a recognition motif for an activity, and modification of
30   said linker moiety by said activity modulates the coupling of said multimerized destabilization domain to said reporter moiety, thereby modulating the stability of said reporter moiety.

57. The method of claim 55, wherein said destabilization domain comprises a ubiquitin homolog.

35   58. The method of claim 57, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases.

59. The method of claim 57, wherein said ubiquitin homolog comprises a mutation at glycine 76.

60. A host cell, comprising a nucleic acid sequence encoding for;

    a) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

    b) a target protein, and

    c) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

        wherein said linker is non-cleavable by a -NH-ubiquitin protein endoproteases.

61. The method of claim 60, wherein said destabilization domain comprises a ubiquitin homolog.

62. The method of claim 61, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases.

63. The method of claim 61, wherein said ubiquitin homolog comprises a mutation at glycine 76.

64. A transgenic animal, comprising a nucleic acid sequence encoding for;

    d) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

    e) a target protein, and

    f) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

        wherein said linker is non-cleavable by a -NH-ubiquitin protein endoproteases.

65. The method of claim 64, wherein said destabilization domain comprises a ubiquitin homolog.

66. The method of claim 65, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases.

67. The method of claim 65, wherein said ubiquitin homolog comprises a mutation at glycine 76.

68. A transgenic plant, comprising a nucleic acid sequence encoding for;

a) a linear multimerized destabilization domain, wherein said multimerized destabilization domain is non-cleavable by a -NH-ubiquitin protein endoproteases, and comprises at least two copies of said destabilization domain,

b) a target protein, and

5   c) a linker moiety that operatively couples said multimerized destabilization domain to said reporter moiety,

wherein said linker is non-cleavable by a -NH-ubiquitin protein endoproteases.

69. The method of claim 68, wherein said destabilization domain comprises a ubiquitin homolog.

10

70. The method of claim 69, wherein said ubiquitin homolog comprises a mutation that prevents cleavage by -NH-ubiquitin protein endoproteases.

71. The method of claim 69, wherein said ubiquitin homolog comprises a mutation at glycine 76.

15

72. A method for identifying a modulator of an activity, comprising:

    a) contacting a cell with a test chemical, wherein said cell comprises,

        a) at least one destabilization domain, wherein said destabilization domain is non-cleavable by -NH-ubiquitin protein endoproteases,

20   ii) a reporter moiety, and

        iii) a linker moiety that operatively couples said destabilization domain to said reporter moiety,

wherein said linker moiety comprises a recognition motif for said activity and modification of said linker moiety by said activity modulates the coupling of said destabilization domain to said reporter moiety thereby modulating the stability of said reporter moiety, and

25   wherein said linker moiety is non-cleavable by said -NH-ubiquitin protein endoproteases,

    b) detecting said reporter moiety, or a product of said reporter moiety in the presence of said test chemical, and

    c) comparing said reporter moiety activity from step b) to the reporter moiety activity in a control cell in the absence of said test chemical.

30   73. The method of claim 72, further comprising the step of contacting said cell with an activator of said activity prior to the addition said test chemical.

74. The method of claim 72, further comprising the step of detecting the viability of said cell

35   75. The method of claim 72, wherein said activity is selected from the group consisting of a protease activity, a protein kinase activity and a phosphoprotein phosphatase activity.

76. The method of claim 72, wherein said reporter moiety is selected from the group consisting of a naturally fluorescent protein homolog, a -lactamase homolog, a -galactosidase homolog, an alkaline phosphatase homolog, a CAT homolog, and a luciferase homolog.

5   77. A test chemical identified by any of the methods of claims 72, 73, 74, 75 or 76.

     78. A pharmaceutical composition comprising a test chemical identified by any one of the methods of claims 72, 73, 74, 75 or 76.

10   79. The pharmaceutical composition of claim 78, further comprising a pharaceutically acceptable carrier.

15

Figure 1

Figure 2

A.

WT Bla

1XUb-Bla

2XUb-Bla

3XUb-Bla

4XUb-Bla

0   5   10   20   30   45   60

B.

UbG76V-Bla fusions

WT Bla   1XUb   2XUb   3XUb   4XUb

MG132:   − +   − +   − +   − +   − +
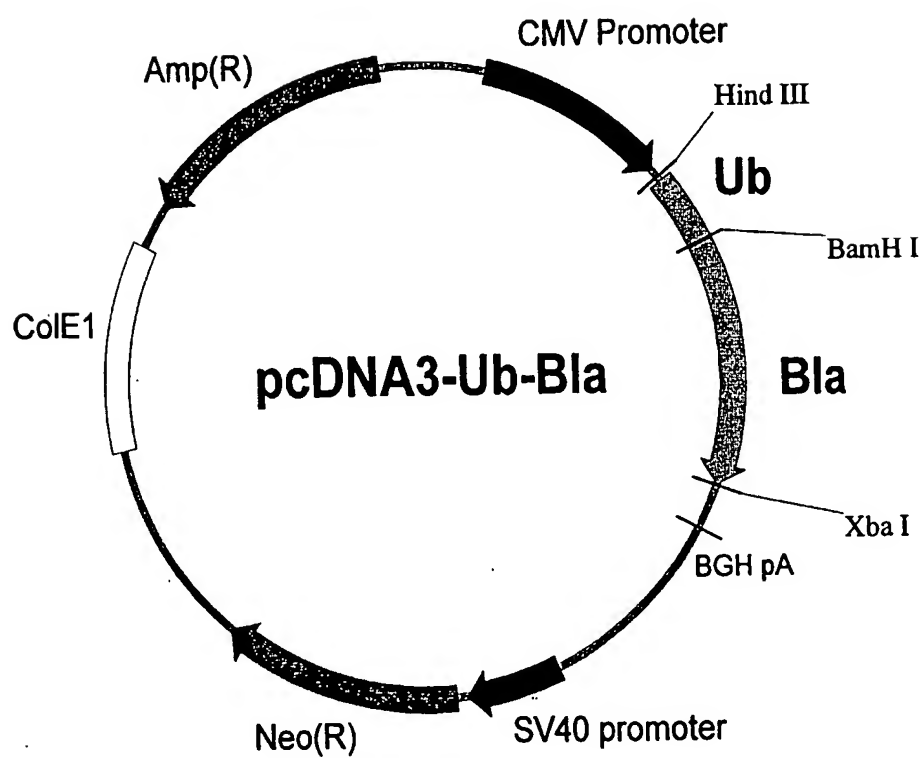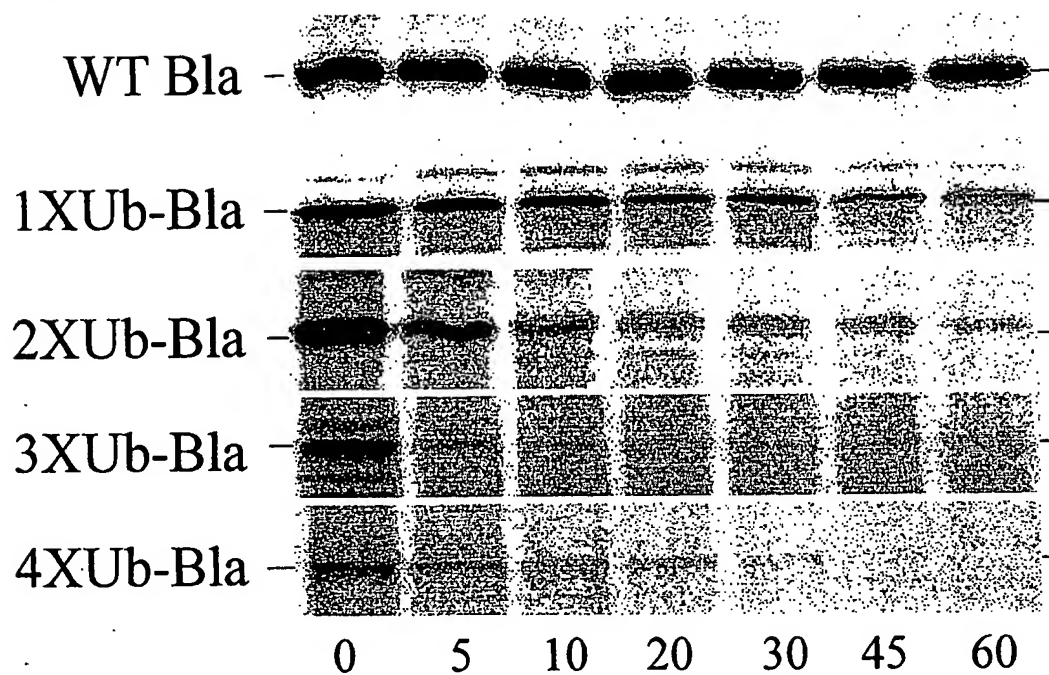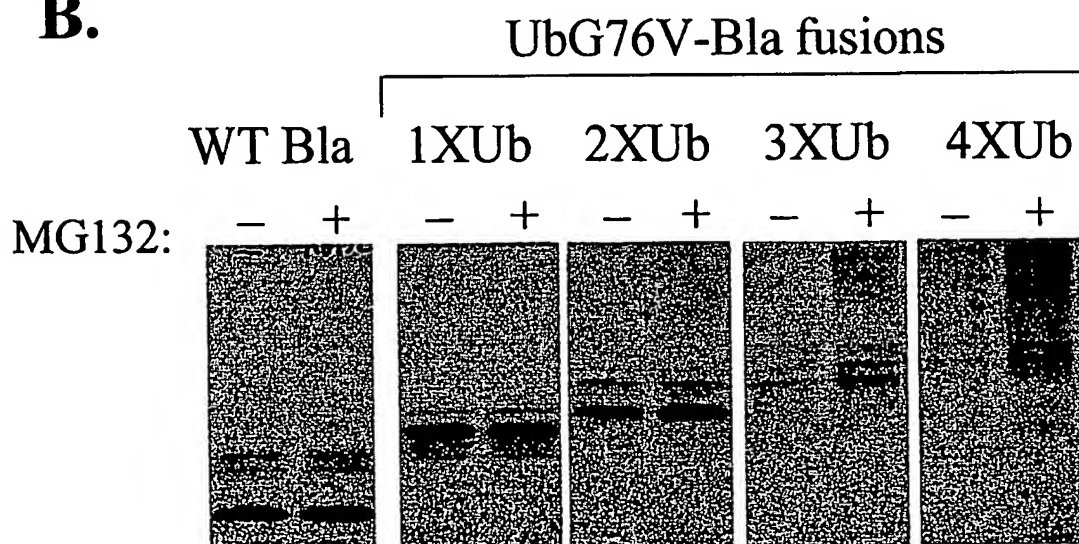
Figure 3

Figure 4

Figure 5

WT Bla

1XUb-Bla

2XUb-Bla

3XUb-Bla

4XUb-Bla

- ■ WT Bla
- □ 1X Ub-Bla
- ▨ 2X Ub-Bla
- ▨ 3X Ub-Bla
- ▨ 4X Ub-Bla

%Bla+ Cells

Figure 6

Figure 7

A.

B.
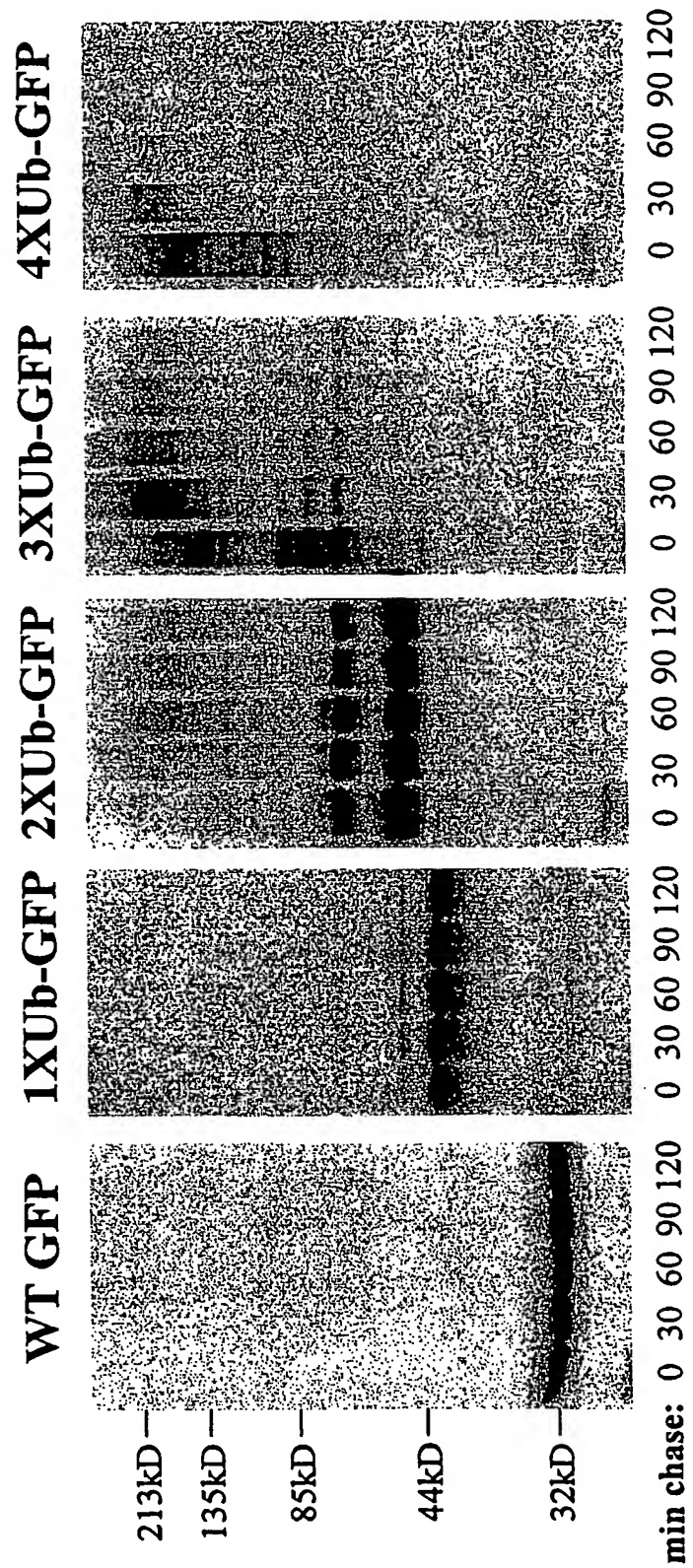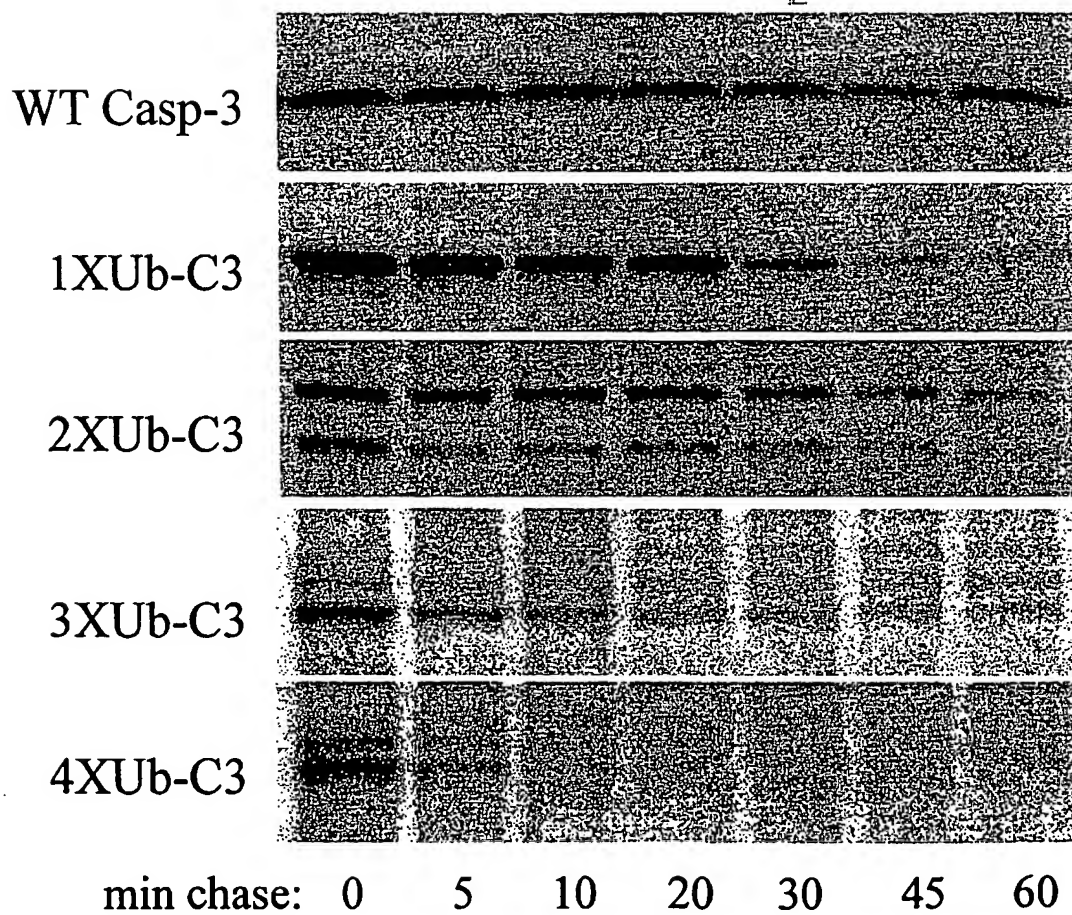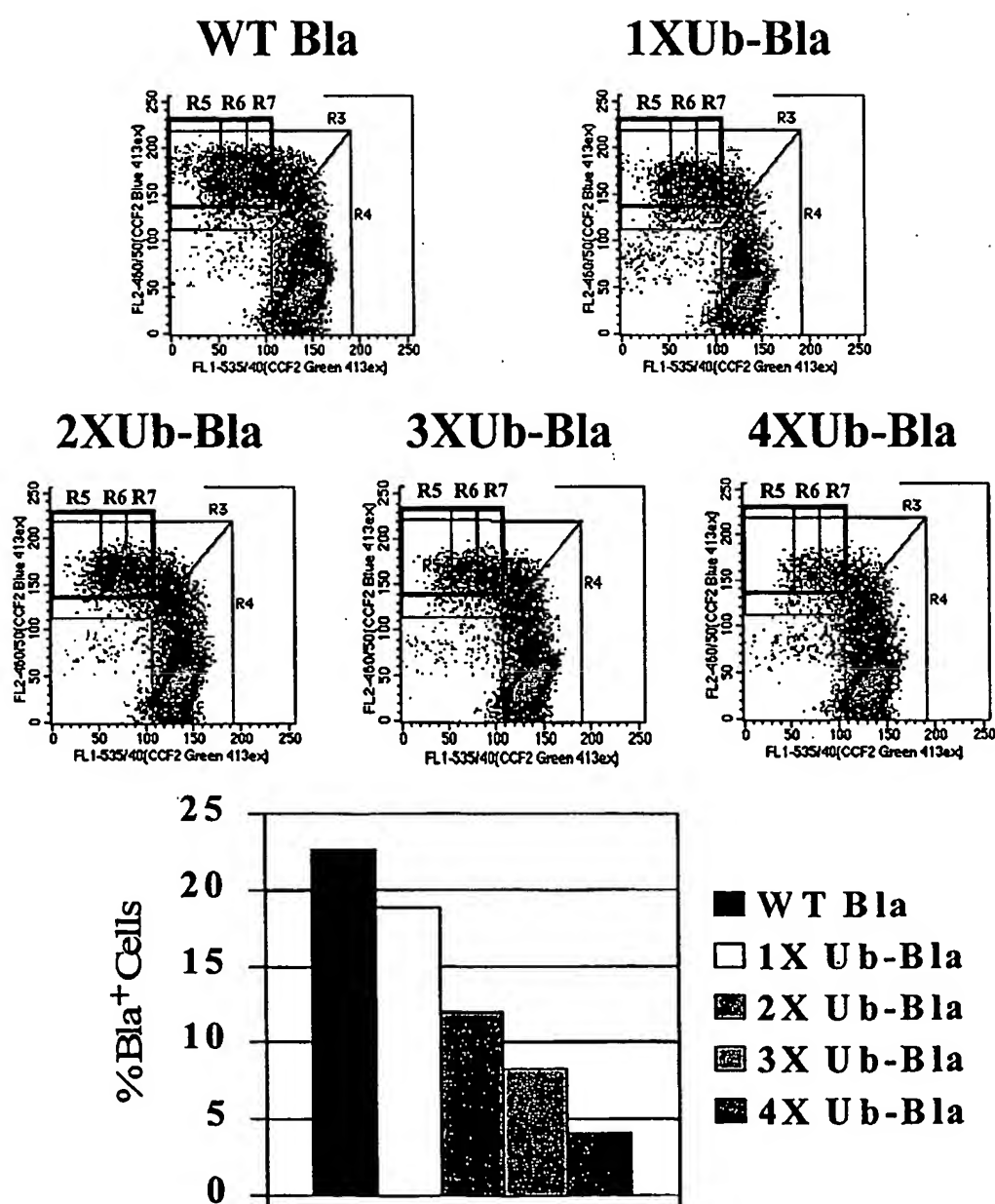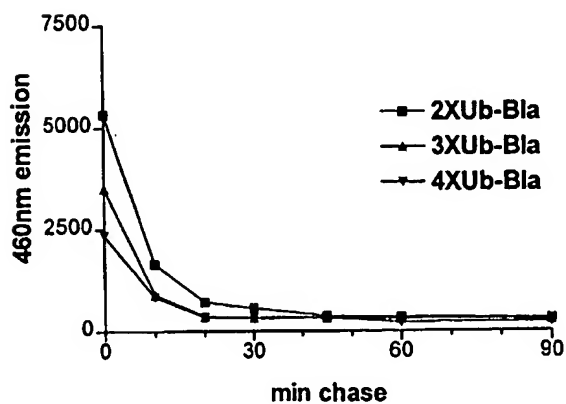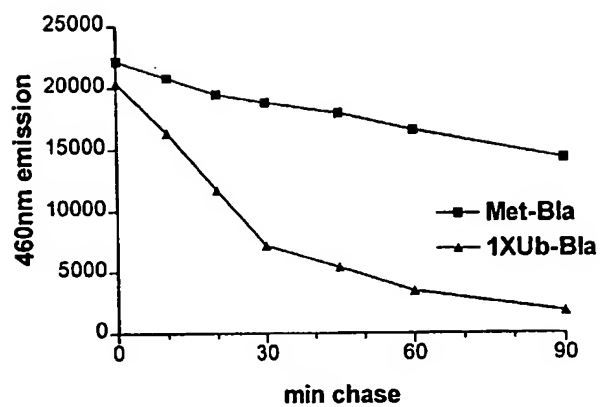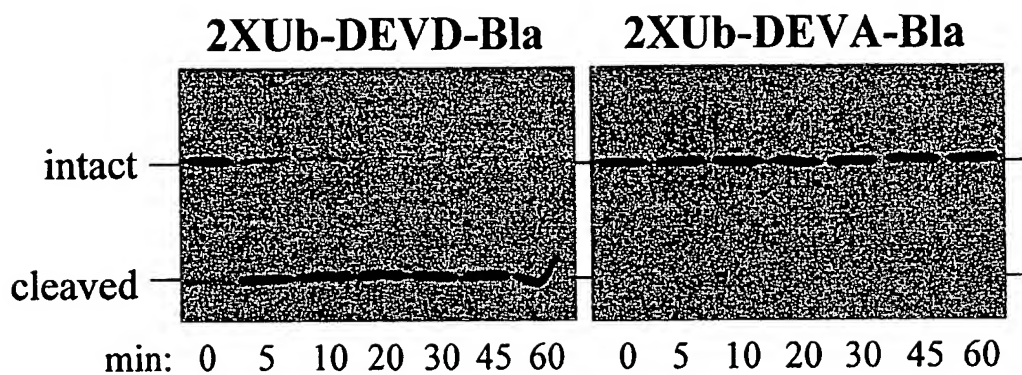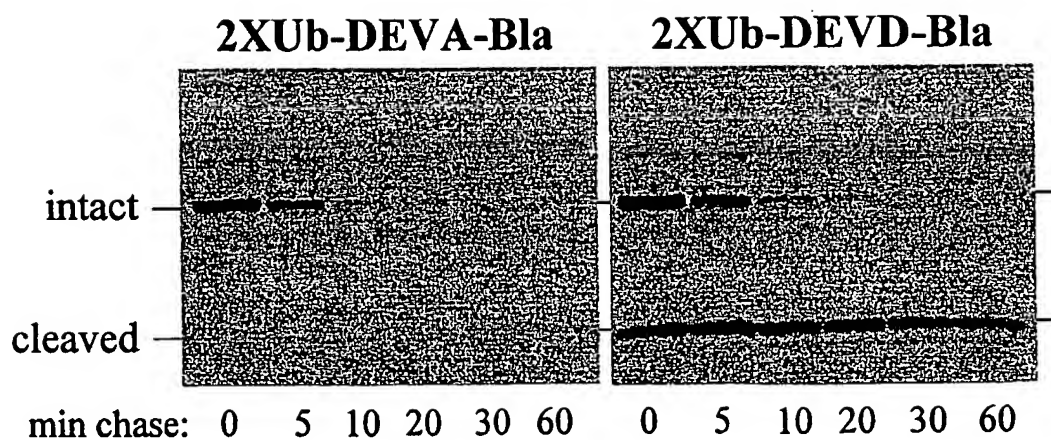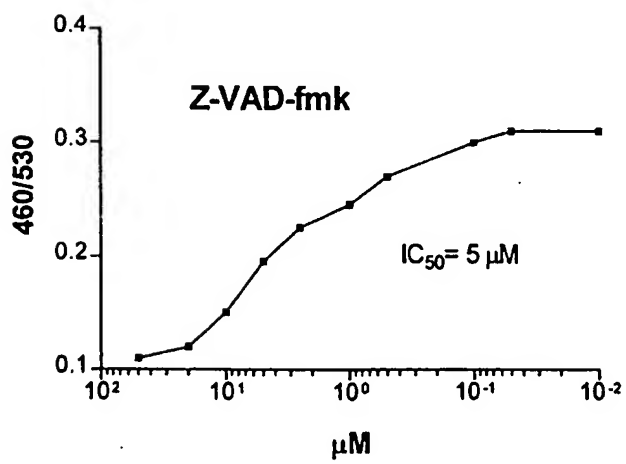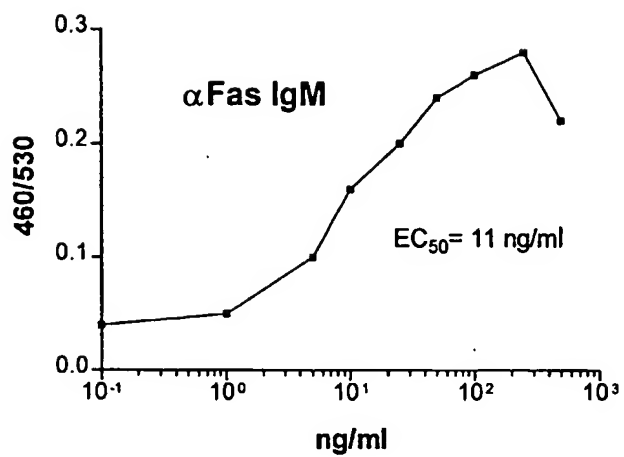
Figure 8

Figure 9

A.

HRV16-Bla

B.

3XUb-HRV14-Bla

Figure 10

0 min   20 min chase
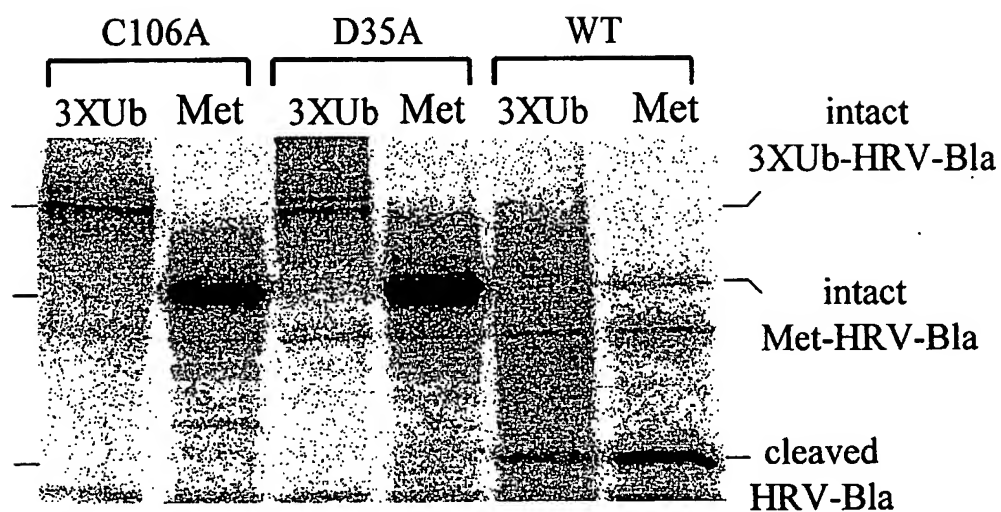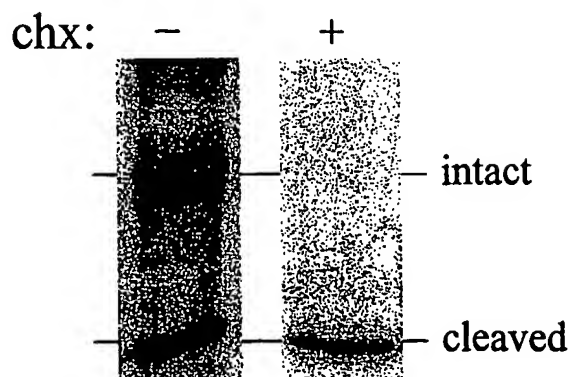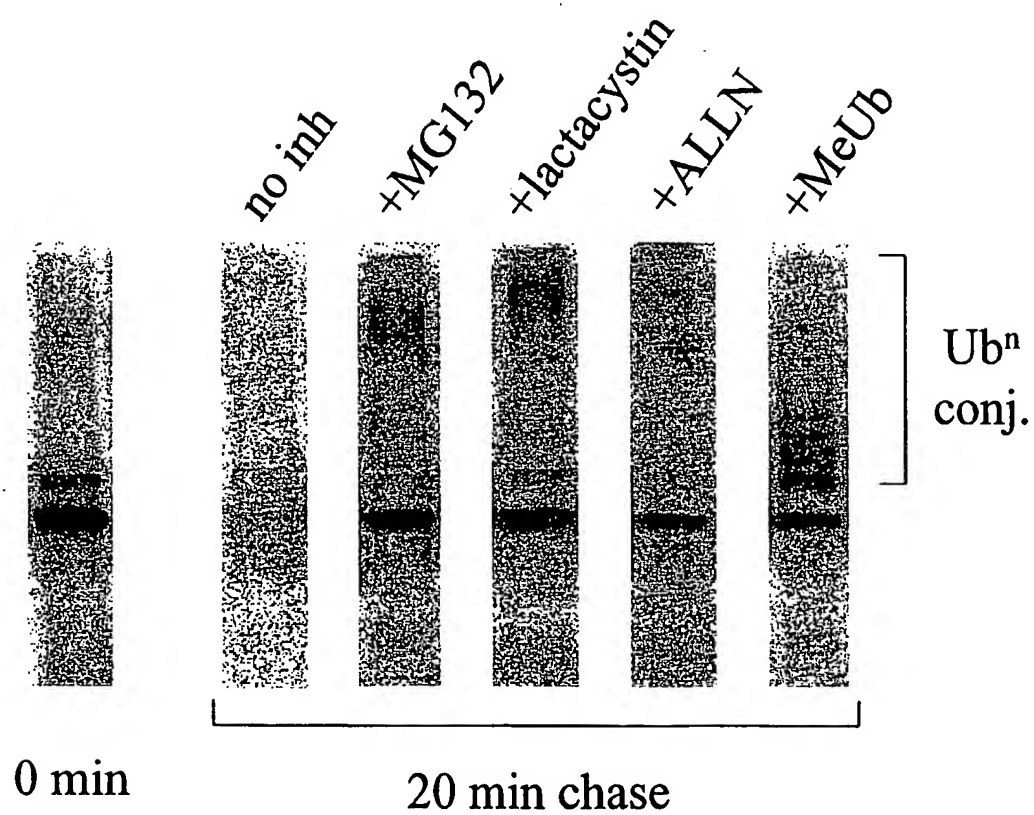
Figure 11
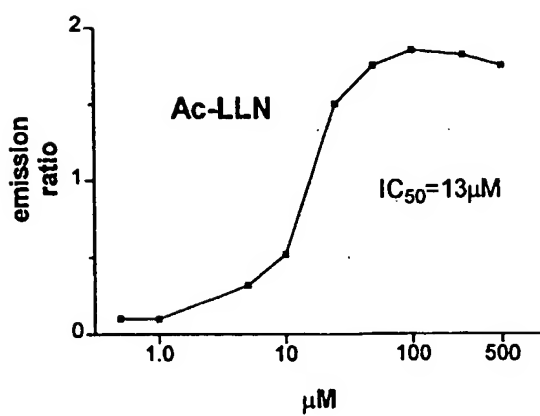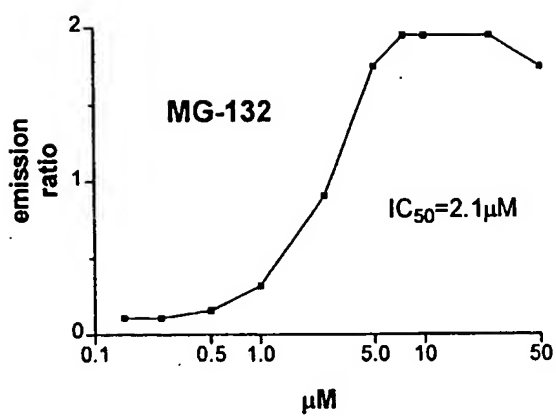
SEQUENCE ID. LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 72

(2) INFORMATION FOR SEQ. ID. NO. 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: peptide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(B) LOCATION: 1....6

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.:1:

DSGLDS

(2) INFORMATION FOR SEQ. ID. NO.: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 228 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....228

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 2:

ATG GAG ATC TTC GTG AAG ACT CTG ACT GGT AAG ACC ATC ACC CTC GAA GTG GAG CCG AGT GAC ACC ATT GAG

AAT GTC AAG GCA AAG ATC CAA GAC AAG GAA GGC ATC CCT CCT GAC CAG CAG AGG TTG ATC TTT GCT GGG AAA

CAG CTG GAA GAT GGA CGC ACC CTG TCT GAC TAC AAC ATC CAG AAA GAG TCC ACC CTG CAC CTG GTA CTC CGT

CTC AGA GGT GGG

5

(2) INFORMATION FOR SEQ. ID. NO.:3 (BLA):

(i) SEQUENCE CHARACTERISTICS:

10   (A) LENGTH: 795 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

20   (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....795

25

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 3:

range 1 to 795

30
```
          10      20      30      40      50
          .   .   .   .   .   .   .   .   .   .
      ATG AGT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG
      Met Ser His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu
35
          60      70      80      90      100
          .   .   .   .   .   .   .   .   .   .
      GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT
      Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu
40
          110     120     130     140     150
          .   .   .   .   .   .   .   .   .   .
      GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT
      Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val
45
          160     170     180     190     200
          .   .   .   .   .   .   .   .   .   .
      CTG CTA TGT GGC GCG GTA TTA TCC CGT GTT GAC GCC GGG CAA GAG CAA CTC
      Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly Gln Glu Gln Leu
50

          210     220     230     240     250
          .   .   .   .   .
```

2

· · · · · · · · · · ·
GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC
Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val

5   260   270   280   290   300
       · ·   · ·   · ·   · ·   · ·
ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT
Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala

10   310   320   330   340   350
       · ·   · ·   · ·   · ·   · ·
GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC
Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile

15   360   370   380   390   400
       · ·   · ·   · ·   · ·   · ·
GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA
Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val

20   410   420   430   440   450
       · ·   · ·   · ·   · ·   · ·
ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC
Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp

25   460   470   480   490   500   510
       · ·   · ·   · ·   · ·   · ·   ·
GAG CGT GAC ACC ACG ATG CCT GCA GCA ATG GCA ACA ACG TTG CGC AAA CTA
Glu Arg Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr Leu Arg Lys Leu

30   520   530   540   550   560
       · ·   · ·   · ·   · ·   · ·
TTA ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG
Leu Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp

35   570   580   590   600   610
       · ·   · ·   · ·   · ·   · ·
ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala

40   620   630   640   650   660
       · ·   · ·   · ·   · ·   · ·
GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT
Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly

45   670   680   690   700   710
       · ·   · ·   · ·   · ·   · ·
ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC
Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile

50   720   730   740   750   760
       · ·   · ·   · ·   · ·   · · ·
TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT
Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala

55   770   780   790

3

• • • • • •

GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
Glu Ile Gly Ala Ser Leu Ile Lys His Trp

5

(2) INFORMATION FOR SEQ. ID. NO.:4:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 858 base pairs

(B) TYPE: nucleic acid

15   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

20

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

25   (B) LOCATION: 1.....858

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO. 4:

range 1 to 858

30
```
        10      20      30      40      50
    •  •  •  •  •  •  •  •  •  •
    ATG AGA ATT CAA CAT TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG GCA TTT
    Met Arg Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala Phe
```

35
```
        60      70      80      90      100
    •  •  •  •  •  •  •  •  •  •
    TGC CTT CCT GTT TTT GGT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT
    Cys Leu Pro Val Phe Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala
```

40
```
        110     120     130     140     150
    •  •  •  •  •  •  •  •  •  •
    GAA GAT CAG TTG GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC
    Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser
```

45
```
        160     170     180     190     200
    •  •  •  •  •  •  •  •  •  •
    GGT AAG ATC CTT GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC
    Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser
```

50
```
        210     220     230     240     250
    •  •  •  •  •  •  •  •  •  •
    ACT TTT AAA GTT CTG CTA TGT GGC GCG GTA TTA TCC CGT GTT GAC GCC GGG
    Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly
```

55

4

```
         260     270     280     290     300
          •   •   •   •   •   •   •   •   •   •
      CAA GAG CAA CTC GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG
      Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu
5


         310     320     330     340     350
          •   •   •   •   •   •   •   •   •   •
10    TAC TCA CCA GTC ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA
      Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu


         360     370     380     390     400
          •   •   •   •   •   •   •   •   •   •
15    TTA TGC AGT GCT GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT
      Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu


         410     420     430     440     450
          •   •   •   •   •   •   •   •   •   •
20    CTG ACA ACG ATC GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG
      Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met


         460     470     480     490     500     510
          •   •   •   •   •   •   •   •   •   •   •
25    GGG GAT CAT GTA ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC
      Gly Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala


           520     530     540     550     560
          •   •   •   •   •   •   •   •   •   •
30    ATA CCA AAC GAC GAG CGT GAC ACC ACG ATG CCT GCA GCA ATG GCA ACA ACG
      Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr


           570     580     590     600     610
          •   •   •   •   •   •   •   •   •   •
35    TTG CGC AAA CTA TTA ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA
      Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln


           620     630     640     650     660
          •   •   •   •   •   •   •   •   •   •
40    TTA ATA GAC TGG ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG
      Leu Ile Asp Trp Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser


           670     680     690     700     710
          •   •   •   •   •   •   •   •   •   •
45    GCC CTT CCG GCT GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT
      Ala Leu Pro Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg


           720     730     740     750     760
          •   •   •   •   •   •   •   •   •   •
50    GGG TCT CGC GGT ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT
      Gly Ser Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg


         770     780     790     800     810
          •   •   •   •   •   •   •   •   •   •
55    ATC GTA GTT ATC TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT
```

5

Ile Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn

```
       820     830     840     850
        •   •   •   •   •   •   •   •
```

AGA CAG ATC GCT GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp

(2) INFORMATION FOR SEQ. ID. NO.: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 795 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....795

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 5:

range 1 to 795

```
       10      20      30      40      50
    •   •   •   •   •   •   •   •   •
```

ATG GGG CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG
Met Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu

```
       60      70      80      90      100
    •   •   •   •   •   •   •   •   •
```

GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT
Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu

```
           110     120     130     140     150
    •   •   •   •   •   •   •   •   •   •
```

GAG AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT
Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val

```
          160     170     180     190     200     210
    •   •   •   •   •   •   •   •   •   •   •
```

CTG CTA TGT GGC GCG GTA TTA TCC CGT GAT GAC GCC GGG CAA GAG CAA CTC
Leu Leu Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu

```
       220     230     240     250     260
    •   •   •   •   •   •   •   •   •
```

GGT CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC
Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val

```
           270     280     290     300     310
5       •   •   •   •   •   •   •   •   •
```
ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT
Thr Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala

```
10        320     330     340     350     360
        •   •   •   •   •   •   •   •   •
```
GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC
Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile

```
15        370     380     390     400     410
        •   •   •   •   •   •   •   •   •
```
GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA
Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val

```
20        420     430     440     450     460
        •   •   •   •   •   •   •   •   •   •
```
ACT CGC CTT GAT CAT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC
Thr Arg Leu Asp His Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp

```
25        470     480     490     500     510
        •   •   •   •   •   •   •   •   •
```
GAG CGT GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG TTG CGC AAA CTA
Glu Arg Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu

```
30     520     530     540     550     560
     •   •   •   •   •   •   •   •   •
```
TTA ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG
Leu Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp

```
35     570     580     590     600     610
     •   •   •   •   •   •   •   •   •
```
ATG GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala

```
40     620     630     640     650     660
     •   •   •   •   •   •   •   •   •
```
GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT
Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly

```
45     670     680     690     700     710     720
     •   •   •   •   •   •   •   •   •   •   •
```
ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC
Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile

```
50        730     740     750     760     770
        •   •   •   •   •   •   •   •   •
```
TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT
Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala

```
55        780     790
```

. . . . . .

GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG
Glu Ile Gly Ala Ser Leu Ile Lys His Trp

(2) INFORMATION FOR SEQ. ID. NO.: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 792 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....792

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 6:

range 1 to 792

```
       10       20       30       40       50
   .   .   .   .   .   .   .   .   .   .
ATG GAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT CAG TTG GGT
Met Asp Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu Gly

       60       70       80       90      100
   .   .   .   .   .   .   .   .   .   .
GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT GAG
Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu Glu

      110      120      130      140      150
   .   .   .   .   .   .   .   .   .   .
AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT CTG
Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val Leu

      160      170      180      190      200
   .   .   .   .   .   .   .   .   .   .
CTA TGT GGC GCG GTA TTA TCC CGT ATT GAC GCC GGG CAA GAG CAA CTC GGT
Leu Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu Gly

      210      220      230      240      250
   .   .   .   .   .   .   .   .   .   .
CGC CGC ATA CAC TAT TCT CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA
Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr
```

8

```
        260    270    280    290    300
         .  .  .  .  .  .  .  .  .  .

GAA AAG CAT CTT ACG GAT GGC ATG ACA GTA AGA GAA TTA TGC AGT GCT GCC
Glu Lys His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala


        310    320    330    340    350
         .  .  .  .  .  .  .  .  .  .

ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT CTG ACA ACG ATC GGA
Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly


        360    370    380    390    400
         .  .  .  .  .  .  .  .  .  .

GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT GTA ACT
Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr


        410    420    430    440    450
         .  .  .  .  .  .  .  .  .  .

CGC CTT GAT CAT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG
Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu


        460    470    480    490    500    510
         .  .  .  .  .  .  .  .  .  .  .

CGT GAC ACC ACG ATG CCT GTA GCA ATG GCA ACA ACG TTG CGC AAA CTA TTA
Arg Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu


           520    530    540    550    560
            .  .  .  .  .  .  .  .  .  .

ACT GGC GAA CTA CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met


           570    580    590    600    610
            .  .  .  .  .  .  .  .  .  .

GAG GCG GAT AAA GTT GCA GGA CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC
Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly


           620    630    640    650    660
            .  .  .  .  .  .  .  .  .  .

TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT GAG CGT GGG TCT CGC GGT ATC
Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile


           670    680    690    700    710
            .  .  .  .  .  .  .  .  .  .

ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC GTA GTT ATC TAC
Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr


           720    730    740    750    760
            .  .  .  .  .  .  .  .  .  .

ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT GAG
Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu


           770    780    790
            .  .  .  .  .

ATA GGT GCC TCA CTG ATT AAG CAT TGG
Ile Gly Ala Ser Leu Ile Lys His Trp
```

(2) INFORMATION FOR SEQ. ID. NO.: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 786 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....786

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 7:

range 1 to 786

```
        10      20      30      40      50
   .  .   .   .   .   .   .   .   .   .
ATG AAA GAT GAT TTT GCA AAA CTT GAG GAA CAA TTT GAT GCA AAA CTC GGG
Met Lys Asp Asp Phe Ala Lys Leu Glu Glu Gln Phe Asp Ala Lys Leu Gly   .


        60      70      80      90      100
   .  .   .   .   .   .   .   .   .   .
ATC TTT GCA TTG GAT ACA GGT ACA AAC CGG ACG GTA GCG TAT CGG CCG GAT
Ile Phe Ala Leu Asp Thr Gly Thr Asn Arg Thr Val Ala Tyr Arg Pro Asp


        110     120     130     140     150
   .  .   .   .   .   .   .   .   .   .
GAG CGT TTT GCT TTT GCT TCG ACG ATT AAG GCT TTA ACT GTA GGC GTG CTT
Glu Arg Phe Ala Phe Ala Ser Thr Ile Lys Ala Leu Thr Val Gly Val Leu


        160     170     180     190     200
   .  .   .   .   .   .   .   .   .   .
TTG CAA CAG AAA TCA ATA GAA GAT CTG AAC CAG AGA ATA ACA TAT ACA CGT        Leu Gln Gln Lys Ser Ile Glu Asp
Leu Asn Gln Arg Ile Thr Tyr Thr Arg


       210     220     230     240     250
  .  .   .   .   .   .   .   .   .   .
GAT GAT CTT GTA AAC TAC AAC CCG ATT ACG GAA AAG CAC GTT GAT ACG GGA
Asp Asp Leu Val Asn Tyr Asn Pro Ile Thr Glu Lys His Val Asp Thr Gly


       260     270     280     290     300
  .  .   .   .   .   .   .   .   .   .
ATG ACG CTC AAA GAG CTT GCG GAT GCT TCG CTT CGA TAT AGT GAC AAT GCG
Met Thr Leu Lys Glu Leu Ala Asp Ala Ser Leu Arg Tyr Ser Asp Asn Ala
```

```
            310     320     330     340     350
             .   .   .   .   .   .   .   .   .   .
        GCA CAG AAT CTC ATT CTT AAA CAA ATT GGC GGA CCT GAA AGT TTG AAA AAG
5       Ala Gln Asn Leu Ile Leu Lys Gln Ile Gly Gly Pro Glu Ser Leu Lys Lys


            360     370     380     390     400
             .   .   .   .   .   .   .   .   .   .
10      GAA CTG AGG AAG ATT GGT GAT GAG GTT ACA AAT CCC GAA CGA TTC GAA CCA
        Glu Leu Arg Lys Ile Gly Asp Glu Val Thr Asn Pro Glu Arg Phe Glu Pro


            410     420     430     440     450
             .   .   .   .   .   .   .   .   .   .
15      GAG TTA AAT GAA GTG AAT CCG GGT GAA ACT CAG GAT ACC AGT ACA GCA AGA
        Glu Leu Asn Glu Val Asn Pro Gly Glu Thr Gln Asp Thr Ser Thr Ala Arg


            460     470     480     490     500     510
             .   .   .   .   .   .   .   .   .   .   .
20      GCA CTT GTC ACA AGC CTT CGA GCC TTT GCT CTT GAA GAT AAA CTT CCA AGT
        Ala Leu Val Thr Ser Leu Arg Ala Phe Ala Leu Glu Asp Lys Leu Pro Ser


                520     530     540     550     560
                 .   .   .   .   .   .   .   .   .
25      GAA AAA CGC GAG CTT TTA ATC GAT TGG ATG AAA CGA AAT ACC ACT GGA GAC
        Glu Lys Arg Glu Leu Leu Ile Asp Trp Met Lys Arg Asn Thr Thr Gly Asp


                570     580     590     600     610
                 .   .   .   .   .   .   .   .   .
30      GCC TTA ATC CGT GCC GGA GCG GCA TCA TAT GGA ACC CGG AAT GAC ATT GCC
        Ala Leu Ile Arg Ala Gly Val Pro Asp Gly Trp Glu Val Ala Asp Lys Thr


                620     630     640     650     660
                 .   .   .   .   .   .   .   .   .
35      ATC ATT TGG CCG CCA AAA GGA GAT CCT GTC GGT GTG CCG GAC GGT TGG GAA
        Gly Ala Ala Ser Tyr Lys Gly Asp Pro Val Gly Thr Arg Asn Asp Ile Ala


                670     680     690     700     710
                 .   .   .   .   .   .   .   .   .
40      GTG GCT GAT AAA ACT GTT CTT GCA GTA TTA TCC AGC AGG GAT AAA AAG GAC
        Ile Ile Trp Pro Pro Val Leu Ala Val Leu Ser Ser Arg Asp Lys Lys Asp


                720     730     740     750     760
                 .   .   .   .   .   .   .   .   .   .
45      GCC AAG TAT GAT GAT AAA CTT ATT GCA GAG GCA ACA AAG GTG GTA ATG AAA
        Ala Lys Tyr Asp Asp Lys Leu Ile Ala Glu Ala Thr Lys Val Val Met Lys


                770     780
                 .   .   .   .
50      GCC TTA AAC ATG AAC GGC AAA
        Ala Leu Asn Met Asn Gly Lys
```


        (2) INFORMATION FOR SEQ. ID. NO. 8:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 720 base pairs

5   (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

15   (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....720

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO. 8:

20
ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC

GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC

25   GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC

GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC TTC TCC TAC GGC

GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC

30
AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG

GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC

35   CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC

ATC CTG GGG CAC AAC CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC

ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC

40
AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC

CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC

45   CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG

CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC

AAG TAA
50

(2) INFORMATION FOR SEQ. ID. NO.: 9:

(i) SEQUENCE CHARACTERISTICS:
55

(A) LENGTH: 690 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10   (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

15   (B) LOCATION: 1....690

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.9:

20   ATG GCT CTT TCA AAC AAG TTT ATC GGA GAT GAC ATG AAA ATG ACC TAC CAT

ATG GAT GGC TGT GTC AAT GGG CAT TAC TTT ACC GTC AAA GGT GAA GGC AAC

GGG AAG CCA TAC GAA GGG ACG CAG ACT TCG ACT TTT AAA GTC ACC ATG GCC

25   AAC GGT GGG CCC CTT GCA TTC TCC TTT GAC ATA CTA TCT ACA GTG TTC AAA

TAT GGA AAT CGA TGC TTT ACT GCG TAT CCT ACC AGT ATG CCC GAC TAT TTC

30   AAA CAA GCA TTT CCT GAC GGA ATG TCA TAT GAA AGG ACT TTT ACC TAT GAA

GAT GGA GGA GTT GCT ACA GCC AGT TGG GAA ATA AGC CTT AAA GGC AAC TGC

TTT GAG CAC AAA TCC ACG TTT CAT GGA GTG AAC TTT CCT GCT GAT GGA CCT

35   GTG ATG GCG AAG AAG ACA ACT GGT TGG GAC CCA TCT TTT GAG AAA ATG ACT

GTC TGC GAT GGA ATA TTG AAG GGT GAT GTC ACC GCG TTC CTC ATG CTG CAA

40   GGA GGT GGC AAT TAC AGA TGC CAA TTC CAC ACT TCT TAC AAG ACA AAA AAA

CCG GTG ACG ATG CCA CCA AAC CAT GTG GTG GAA CAT CGC ATT GCG AGG ACC

GAC CTT GAC AAA GGT GGC AAC AGT GTT CAG CTG ACG GAG CAC GCT GTT GCA

45   CAT ATA ACC TCT GTT GTC CCT TTC TGA

(2) INFORMATION FOR SEQ. ID. NO.10:

50   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 696 base pairs

55   (B) TYPE: nucleic acid

13

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

10   (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....696

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.10:

15

ATG GCT CAG TCA AAG CAC GGT CTA ACA AAA GAA ATG ACA ATG AAA TAC CGT

ATG GAA GGG TGC GTC GAT GGA CAT AAA TTT GTG ATC ACG GGA GAG GGC ATT

20

GGA TAT CCG TTC AAA GGG AAA CAG GCT ATT AAT CTG TGT GTG GTC GAA GGT

GGA CCA TTG CCA TTT GCC GAA GAC ATA TTG TCA GCT GCC TTT AAC TAC GGA

AAC AGG GTT TTC ACT GAA TAT CCT CAA GAC ATA GTT GAC TAT TTC AAG AAC

25

TCG TGT CCT GCT GGA TAT ACA TGG GAC AGG TCT TTT CTC TTT GAG GAT GGA

GCA GTT TGC ATA TGT AAT GCA GAT ATA ACA GTG AGT GTT GAA GAA AAC TGC

30

ATG TAT CAT GAG TCC AAA TTT TAT GGA GTG AAT TTT CCT GCT GAT GGA CCT

GTG ATG AAA AAG ATG ACA GAT AAC TGG GAG CCA TCC TGC GAG AAG ATC ATA

35

CCA GTA CCT AAG CAG GGG ATA TTG AAA GGG GAT GTC TCC ATG TAC CTC CTT

CTG AAG GAT GGT GGG CGT TTA CGG TGC CAA TTC GAC ACA GTT TAC AAA GCA

AAG TCT GTG CCA AGA AAG ATG CCG GAC TGG CAC TTC ATC CAG CAT AAG CTC

40

ACC CGT GAA GAC CGC AGC GAT GCT AAG AAT CAG AAA TGG CAT CTG ACA GAA

CAT GCT ATT GCA TCC GGA TCT GCA TTG CCC TGA

45

(2) INFORMATION FOR SEQ. ID. NO.11:

(i) SEQUENCE CHARACTERISTICS:

50   (A) LENGTH: 696 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55

14

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

5   (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1...698

10
(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.11:

ATG GCT CAT TCA AAG CAC GGT CTA AAA GAA GAA ATG ACA ATG AAA TAC CAC
15
ATG GAA GGG TGC GTC AAC GGA CAT AAA TTT GTG ATC ACG GGC GAA GGC ATT

GGA TAT CCG TTC AAA GGG AAA CAG ACT ATT AAT CTG TGT GTG ATC GAA GGG

20   GGA CCA TTG CCA TTT TCC GAA GAC ATA TTG TCA GCT GGC TTT AAG TAC GGA

GAC AGG ATT TTC ACT GAA TAT CCT CAA GAC ATA GTA GAC TAT TTC AAG AAC

TCG TGT CCT GCT GGA TAT ACA TGG GGC AGG TCT TTT CTC TTT GAG GAT GGA

25
GCA GTC TGC ATA TGC AAT GTA GAT ATA ACA GTG AGT GTC AAA GAA AAC TGC

ATT TAT CAT AAG AGC ATA TTT AAT GGA ATG AAT TTT CCT GCT GAT GGA CCT

30   GTG ATG AAA AAG ATG ACA ACT AAC TGG GAA GCA TCC TGC GAG AAG ATC ATG

CCA GTA CCT AAG CAG GGG ATA CTG AAA GGG GAT GTC TCC ATG TAC CTC CTT

CTG AAG GAT GGT GGG CGT TAC CGG TGC CAG TTC GAC ACA GTT TAC AAA GCA

35
AAG TCT GTG CCA AGT AAG ATG CCG GAG TGG CAC TTC ATC CAG CAT AAG CTC

CTC CGT GAA GAC CGC AGC GAT GCT AAG AAT CAG AAG TGG CAG CTG ACA GAG

40   CAT GCT ATT GCA TTC CCT TCT GCC TTG GCC TGA

(2) INFORMATION FOR SEQ. ID. NO.12:

45   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 699 base pairs

(B) TYPE: nucleic acid

50
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

55   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

5

(B) LOCATION: 1...699

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.12:

10

ATG AGT TGT TCC AAG AGT GTG ATC AAG GAA GAA ATG TTG ATC GAT CTT CAT

CTG GAA GGA ACG TTC AAT GGG CAC TAC TTT GAA ATA AAA GGC AAA GGA AAA

15   GGA CAG CCT AAT GAA GGC ACC AAT ACC GTC ACG CTC GAG GTT ACC AAG GGT

GGA CCT CTG CCA TTT GGT TGG CAT ATT TTG TGC CCA CAA TTT CAG TAT GGA

AAC AAG GCA TTT GTC CAC CAC CCT GAC AAC ATA CAT GAT TAT CTA AAG CTG

20

TCA TTT CCG GAG GGA TAT ACA TGG GAA CGG TCC ATG CAC TTT GAA GAC GGT

GGC TTG TGT TGT ATC ACC AAT GAT ATC AGT TTG ACA GGC AAC TGT TTC TAC

25   TAC GAC ATC AAG TTC ACT GGC TTG AAC TTT CCT CCA AAT GGA CCC GTT GTG

CAG AAG AAG ACA ACT GGC TGG GAA CCG AGC ACT GAG CGT TTG TAT CCT CGT

GAT GGT GTG TTG ATA GGA GAC ATC CAT CAT GCT CTG ACA GTT GAA GGA GGT

30   GGT CAT TAC GCA TGT GAC ATT AAA ACT GTT TAC AGG GCC AAG AAG GCC GCC

TTG AAG ATG CCA GGG TAT CAC TAT GTT GAC ACC AAA CTG GTT ATA TGG AAC

35   AAC GAC AAA GAA TTC ATG AAA GTT GAG GAG CAT GAA ATC GCC GTT GCA CGC

CAC CAT CCG TTC TAT GAG CCA AAG AAG GAT AAG TAA

40   (2) INFORMATION FOR SEQ. ID. NO.13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 678 base pairs

45

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

50   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

55

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....678

5   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.13:

ATG AGG TCT TCC AAG AAT GTT ATC AAG GAG TTC ATG AGG TTT AAG GTT CGC ATG

10   GAA GGA ACG GTC AAT GGG CAC GAG TTT GAA ATA GAA GGC GAA GGA GAG GGG AGG

CCA TAC GAA GGC CAC AAT ACC GTA AAG CTT AAG GTA ACC AAG GGG GGA CCT TTG

CCA TTT GCT TGG GAT ATT TTG TCA CCA CAA TTT CAG TAT GGA AGC AAG GTA TAT

15   GTC AAG CAC CCT GCC GAC ATA CCA GAC TAT AAA AAG CTG TCA TTT CCT GAA GGA

TTT AAA TGG GAA AGG GTC ATG AAC TTT GAA GAC GGT GGC GTC GTT ACT GTA ACC

20   CAG GAT TCC AGT TTG CAG GAT GGC TGT TTC ATC TAC AAG GTC AAG TTC ATT GGC

GTG AAC TTT CCT TCC GAT GGA CCT GTT ATG CAA AAG AAG ACA ATG GGC TGG GAA

GCC AGC ACT GAG CGT TTG TAT CCT CGT GAT GGC GTG TTG AAA GGA GAG ATT CAT

25   AAG GCT CTG AAG CTG AAA GAC GGT GGT CAT TAC CTA GTT GAA TTC AAA AGT ATT

TAC ATG GCA AAG AAG CCT GTG CAG CTA CCA GGG TAC TAC TAT GTT GAC TCC AAA

30   CTG GAT ATA ACA AGC CAC AAC GAA GAC TAT ACA ATC GTT GAG CAG TAT GAA AGA

ACC GAG GGA CGC CAC CAT CTG TTC CTT TAA

35   (2) INFORMATION FOR SEQ. ID. NO.14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 801 base pairs

40
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

45   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

50
(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......801

55   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.14:

17

ATG AAG TGT AAA TTT GTG TTC TGC CTG TCC TTC TTG GTC CTC GCC ATC ACA

AAC GCG AAC ATT TTT TTG AGA AAC GAG GCT GAC TTA GAA GAG AAG ACA TTG

AGA ATA CCA AAA GCT CTA ACC ACC ATG GGT GTG ATT AAA CCA GAC ATG AAG

ATT AAG CTG AAG ATG GAA GGA AAT GTA AAC GGG CAT GCT TTT GTG ATC GAA

GGA GAA GGA GAA GGA AAG CCT TAC GAT GGG ACA CAC ACT TTA AAC CTG GAA

GTG AAG GAA GGT GCG CCT CTG CCT TTT TCT TAC GAT ATC TTG TCA AAC GCG

TTC CAG TAC GGA AAC AGA GCA TTG ACA AAA TAC CCA GAC GAT ATA GCA GAC

TAT TTC AAG CAG TCG TTT CCC GAG GGA TAT TCC TGG GAA AGA ACC ATG ACT

TTT GAA GAC AAA GGC ATT GTC AAA GTG AAA AGT GAC ATA AGC ATG GAG GAA

GAC TCC TTT ATC TAT GAA ATT CGT TTT GAT GGG ATG AAC TTT CCT CCC AAT

GGT CCG GTT ATG CAG AAA AAA ACT TTG AAG TGG GAA CCA TCC ACT GAG ATT

ATG TAC GTG CGT GAT GGA GTG CTG GTC GGA GAT ATT AGC CAT TCT CTG TTG

CTG GAG GGA GGT GGC CAT TAC CGA TGT GAC TTC AAA AGT ATT TAC AAA GCA

AAA AAA GTT GTC AAA TTG CCA GAC TAT CAC TTT GTG GAC CAT CGC ATT GAG

ATC TTG AAC CAT GAC AAG GAT TAC AAC AAA GTA ACG CTG TAT GAG AAT GCA

GTT GCT CGC TAT TCT TTG CTG CCA AGT CAG GCC TAG

(2) INFORMATION FOR SEQ. ID. NO.15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.15:

GATCGGTACCACCATGGAGATCTTCGTGAAGACTCTG

(2) INFORMATION FOR SEQ. ID. NO.16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs

5

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

10   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.16:

15

TGCAGGATCCGTGCATCCCACCTCTGAGACGGAGTACCAG

20

(2) INFORMATION FOR SEQ. ID. NO.17: (UbquitinG76V)

(i) SEQUENCE CHARACTERISTICS:

25   (A) LENGTH: 228 nucleotides

(B) TYPE:

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

35   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.17:
ATG GAG ATC TTC GTG AAG ACT CTG ACT GGT AAG ACC ATC ACC CTC GAA GTG GAG CCG AGT GAC ACC ATT GAG
AAT GTC AAG GCA AAG ATC CAA GAC AAG GAA GGC ATC CCT CCT GAC CAG CAG AGG TTG ATC TTT GCT GGG AAA
CAG CTG GAA GAT GGA CGC ACC CTG TCT GAC TAC AAC ATC CAG AAA GAG TCC ACC CTG CAC CTG GTA CTC CGT
CTC AGA GGT GTG

40

(2) INFORMATION FOR SEQ. ID. NO.18: (Ub5   primer)

(A) LENGTH: 32 nucleotides

45   (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: oligonucleotide

(B) LOCATION: 1....32

55   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.:18:

CGAGATCTACCATGGAAATCTTCGTGAAGACT

5

(2) INFORMATION FOR SEQ. ID. NO.19: (Ub3 primer)

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 22 nucleotides

(B) TYPE: nucleic acid

15   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

20

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

25   (B) LOCATION: 1.....22

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 19:

30   GGATCCGTGGTGCACACCTCTG

(2) INFORMATION FOR SEQ. ID. NO. 20 (BLA5)

(i) SEQUENCE CHARACTERISTICS:

35

(A) LENGTH: 49 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

45   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.20:

GATAGGATCCGGGGCGTGGCTGCACCCAGAAACGCTGGTGAAAGTAAAA

50   (2) INFORMATION FOR SEQ. ID. NO.21: (ABSC107)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

5   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

10
(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.21:
GAACTCTAGATTACCAATGCTTAATCAG

15   (2) INFORMATION FOR SEQ. ID. NO.22: (pcDNA3-Bla)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6180 nucleotides

20
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

25   (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

30
(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6180

35   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:22:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
40   gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccgagctcggat
ccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttg
45   agagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatac
actattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacact
gcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatg
aagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaat
agactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatca
50   ttgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctc
actgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccc
ctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggg
gtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggg
gtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcc
55   cttccttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg

tgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccct
atctcggtctattcttttgatttataaggggattttgggggatttcggcctattggttaaaaaatgagctgattaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagt
tagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcag
aagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctga

5   ctaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagctt
gtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggct
atgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaa
ctgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaa
gtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcga

10   ccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgtt
cgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgact
gtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatc
gccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacga
gatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccac

15   cccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaa
tgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgag
ccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcat
taatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagc
tcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttg

20   ctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctg
gaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcag
ttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacg
acttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggac
agtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcag

25   cagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatc
aaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacct
atctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcga
gacccacgctcaccggctccagatttatcagcaataaaccagccagccgcggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgtt
gccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccgg

30   ttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcat
ggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagt
tgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgct
gttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaa
agggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcaggggttattgtctcatgagcggatacatatttgaatgtatttag

35   aaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

22

(B) LOCATION: 1.....6411

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO: 23:

5   range 1 to 6411

```
          10        20        30        40        50
          .    .    .    .    .    .    .    .    .    .

     gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
10   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
     acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
     gctgaccgcccaacgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact
     gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
     cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
15   accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
     gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
     tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaagtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
     agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgca
     ccacggatccgggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaa
20   gatccttgagagtttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcg
     ccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacgggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtg
     ataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgga
     gctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggca
     acaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctc
25   gcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagat
     aggtgcctcactgattaagcattggtaatctagagggcccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgt
     tgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaataaaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgg
     ggggtggggtggggcaggacagcaagggggaggattgggaagacaataagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctgggg
     ctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttc
30   gctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaact
     tgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaaca
     ctcaaccctatctcggtctattctttgatttataagggattttgggggatttcggctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatg
     tgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctcccca
     gcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgcccc
35   atggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctccc
     gggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggc
     tattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgcc
     ctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctat
40   tgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgc
     ccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagc
     cgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggat
     tcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctt
     tacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctg
     ccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttc
45   ttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttcactgcattctagttgtggtttgtccaa
     actcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacaca
     acatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgc
     cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagc
     ggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag
50   gccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcg
     tttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgta
     ggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg
     taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacac
     tagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtt
55   tgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcat
```

23

gagattatcaaaaaggatcttcacctagatcctttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagt
gaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatg
ataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtct
attaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcatt
5   cagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtt
atcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc
gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggat
cttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttacttttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgc
cgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaa
10   tgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO.24: (pcDNA3-2XUb-Bla)

15   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6678 nucleotides

(B) TYPE: nucleic acid
20
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

25   (ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence
30
(B) LOCATION: 1.....6678

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 24:

35   range 1 to 6678

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
40   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaataacaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggctttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
45   accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
50   gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttg
ggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggc
gcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatgg
catgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttttgcac
55   aacatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttg

24

cgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctgg
ctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactgggggccagatggtaagccctcccgtatcgtagttatctacacgacgggga
gtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgctagagc
tcgctgatcagcctcgactgtgccttctagttgccagccatctgttgttgccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatg
aggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatg
cggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgca
gcgtgaccgctacacttgccagcgcccctagcgcccgctccttcgctttcttcccttccttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatcccttt
agggttccgatttagtgcttacggcacctcgacccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttttcgcccctttgacgttgga
gtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttggggatttcggcctattggttaaaaaa
tgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatc
tcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgc
ccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagta
gtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaaca
agatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcg
cagggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcag
ctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatcctcctgtcatctcaccttgctcctgccgagaaagtatccatcatg
gctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcg
atcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgat
gcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctaccgtgatatt
gctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcg
ggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggac
gccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgttttattgcagcttataatggttacaaataaagcaatagcatcacaaatttc
acaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtca
tagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaatt
gcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc
ctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaa
catgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaa
gtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc
ctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagccc
gaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtag
gcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagc
tcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacgggg
tctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaa
gtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagata
actacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggcc
gagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattg
ctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagct
ccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtg
actggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagt
gctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttact
ttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattg
aagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....6981

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 25:

range 1 to 6981

```
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccggggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccggggcgtggctgcacccagaaacgctggtgaaagta
aaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcac
ttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcac
agaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaa
ggagctaaccgcttttttgcacaacatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcct
gtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttct
gcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgt
agttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggcccctattctatag
tgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccca
ctgtccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggggaggattgggaagacaat
agcaggcatgctggggatgcggtgggctctatggcttcgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcg
ggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctc
taaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttt
ttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttggggattt
cggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcaga
agtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccata
gtcccgcccctaactccgcccatcccgccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcc
tctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgagga
tcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgc
cgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacg
acgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgc
cgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcg
gatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatct
cgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatag
cgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttct
tgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggctt
cggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaag
caatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagag
```

cttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagt
gagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgta
ttgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggg
gataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatca

caaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgctt
accggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacg
aaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattag
cagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcg
gaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcc
tttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatga
agttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctg
actccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaac
cagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagttt
gcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttg
tgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccat
ccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataaccgcgcca
catagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaa
ctgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcata
ctcttcctttttcaatattattgaagcattatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccg
aaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO.: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7164 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1.....7164

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.: 26:

range 1 to 7164

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacgggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgtcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccggggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc

tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagac
catcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaaca
gctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccggggcgtggctgcaccc
agaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaa
cgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttgg
ttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgaca
acgatcggaggaccgaaggagctaaccgctttttttgcacaacatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgag
cgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggata
aagttgcaggaccacttctgcgctcggcccttccggctggctggttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagat
ggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaat
ctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgac
cctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtgggcaggacagcaagg
gggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgta
gcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttccttttctcgccacgttcg
ccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggc
catcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattctttgatt
tataagggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtcccca
ggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatct
caattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagag
gccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatc
aagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaat
cggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcgg
ctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgt
catctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatc
gagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatg
cccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcgga
ccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgca
tcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgcctt
ctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttat
aatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtatacc
gtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagc
ctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcgg
ggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacgg
ttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccg
cccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcc
tgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcca
agctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagc
cactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgc
tgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaa
ggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatc
cttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcg
ttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccag
atttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagt
tcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagtta
catgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattc
tcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgg
gataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacc
cactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaa
atgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttcc
gcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO.27: (pcDNA3-Ub-Met-Bla)

5   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6411 base pairs

(B) TYPE: nucleic acid

10   (C) STRANDEDNESS: double

(D) TOPOLOGY: circular

15   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20   (B) LOCATION: 1......6411

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:27:

25   gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt

agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggggttaggcgtttgcgctgcttcgcgatgtacgggccagatat

acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg

gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataagggactttccattgacgtcaatgggtggactatttacggtaaact

gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt

30   cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc

accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg

gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga

tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc

agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgggat

35   gcacggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaa

gatccttgagagttttcgccccgaagaacgttttccaatgatgagcactttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcg

ccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtg

ataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgga

gctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggca

40   acaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctc

gcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagat

aggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgt

tgtttgcccctcccccgtgccttccttgacctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgg

ggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctgggg

45   ctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttc

gctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgacccccaaaaaact

tgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaaca

ctcaaccctatctcggtctattctttgatttataagggattttggggattcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatg

tgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctcccca

50   gcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgcccc

atggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctccc

gggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggc

tattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgcc

ctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctat

55   tgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgc

ccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagc
cgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggat
tcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctt
tacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctg
ccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgtttccgggacgccggctggatgatcctccagcgcgggggatctcatgctggagttc
ttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaa
actcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacaca
acatacgagccggaagcataaagtgtaaagcctgggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgc
cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagc
ggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag
gccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcg
tttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgta
ggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg
taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacac
tagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtt
tgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcat
gagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagt
gaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatg
ataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtct
attaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcatt
cagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtt
atcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc
gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggat
cttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgc
cgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaa
tgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO.28: (Emerald)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 720 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......720

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:28:
atggtgagcaagggcgaggagctgttcaccgggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcga
tgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccacctcgtgaccaccttcacctacggcgtgcagtgcttcgcccgc
taccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgcc
gaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacag
ccacaaggtctatatcaccgccgacaagcagaagaacggcatcaaggtgaacttcaagacccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccag
cagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagacccccaacgagaagcgcgatcacatggtcct
gctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaa

(2) INFORMATION FOR SEQ. ID. NO.29: (GFP 5' primer)

5   (i) SEQUENCE CHARACTERISTICS:

       (A) LENGTH: 24 base pairs

       (B) TYPE: nucleic acid

10
       (C) STRANDEDNESS: single

       (D) TOPOLOGY: linear

15   (ii) MOLECULE TYPE: oligonucleotide

       (ix) FEATURE:

       (A) NAME / KEY: Coding Sequence

20
       (B) LOCATION: 1......24

       (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:29:
       ggatccgaattcgccaccatggtg

25

30

       (2) INFORMATION FOR SEQ. ID. NO:30:

35
       (i) SEQUENCE CHARACTERISTICS:

       (A) LENGTH: 24 base pairs

40   (B) TYPE: nucleic acid

       (C) STRANDEDNESS: single

       (D) TOPOLOGY: linear

45
       (ii) MOLECULE TYPE: oligonucleotide

       (ix) FEATURE:

50   (A) NAME / KEY: Coding Sequence

       (B) LOCATION: 1......24

       (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:30:
55   ccggaatcaaagcgcttctcagacttactt

2) INFORMATION FOR SEQ. ID. NO.31: (pcDNA3-1XUb-GFP)

5   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6340 base pairs

(B) TYPE: nucleic acid

10   (C) STRANDEDNESS: double

(D) TOPOLOGY: circular

15   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20   (B) LOCATION: 1......6340

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:31:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt

25   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat

acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg

gctgaccgcccaacgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact

gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt

cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc

30   accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg

gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactataggggagacccaagcttggtaccaccatggaga

tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc

agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgca

ccacggatccgaattcgccaccatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgt

35   gtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccttcacccta

cggcgtgcagtgcttcgcccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacga

cggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcac

aagctggagtacaactacaacagccacaaggtctatatcaccgccgacaagcagaagaacggcatcaaggtgaacttcaagacccgccacaacatcgaggacggcagcg

tgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaac

40   gagaagcgcgatcacatggtcctgctggagttcgtgaccgccgcgggatcactctcggcatggacgagctgtacaagtaagtctagagggcccttattctatagtgtcacct

aaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctt

tcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtgggcaggacagcaaggggaggattgggaagacaatagcaggc

atgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtgg

tggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcg

45   gggcatccctttagggttccgatttagtgctttacggcacctcgacccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccct

ttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaacccctatctcggtctattctttgatttataagggattttggggatttcggcctat

tggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgca

aagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcc

cctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagcta

50   ttccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgc

atgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccg

gctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgtt

ccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagt

atccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagc

55   cggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgac

ccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggcta
cccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagtt
cttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgt
tttccgggacgccggctggatgatcctccagcgcgggggatctcatgctggagttcttcgcccacccccaacttgtttattgcagccttataatggttacaaataaagcaatagcat
cacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgta
atcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaact
cacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgct
cttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgca
ggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcg
acgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatac
ctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccg
ttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgag
gtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagag
ttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatctttt
ctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttaaattaaaaatgaagttttaaat
caatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtc
gtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagcc
ggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgtt
gttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaa
gcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagat
gcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcaga
actttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttca
gcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttt
tcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgcca
cctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6607 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6607

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:32:
gacggatcgggagatctcccgatccccttatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggggttaggcgtttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact

gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt

cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgggganttccaagtctcc

accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg

gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc

agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag

acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc

tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat

gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag

aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgaattcgccaccatggtgagcaagggcgaggagctgttcaccgggggtggtgcccatc

ctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccgg

caagctgcccgtgccctggcccaccctcgtgaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatg

cccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagct

gaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaaggtctatatcaccgccgacaagcagaagaacggcatc

aaggtgaacttcaagacccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgaca

accactacctgagcacccagtccgccctgagcaaagacccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggac

gagctgtacaagtaagtctagagggcccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccc ctc

cccgtgccttcttgaccctggaaggtgccactcccactgtccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtgggggtg

gggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggta

tccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccctt

cctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatcccctttagggttccgatttagtgctttacggcacctcgacccccaaaaaacttgattagggtga

tggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatc

tcggtctattcttttgatttataagggattttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttag

ggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaag

tatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaa

tttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtat

atccatttttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatg

actgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactg

caggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtg

ccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgacca

ccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgc

caggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtg

gccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgcc

gctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagat

ttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccacccc

aacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgta

tcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccg

gaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaa

tgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctca

ctcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctg

gcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctgga

agctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagtt

cggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgac

ttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacag

tatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagca

gattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaa

aaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctat

ctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgaga

cccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgcc

gggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc

ccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggt

tatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgc

tcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgtt

gagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaag

ggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaa

aaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc.

5   (2) INFORMATION FOR SEQ. ID. NO:33:

     (i) SEQUENCE CHARACTERISTICS:

     (A) LENGTH: 6850 base pairs

10   (B) TYPE: nucleic acid

     (C) STRANDEDNESS: double

15   (D) TOPOLOGY: circular

     (ii) MOLECULE TYPE: cDNA

     (ix) FEATURE:

20   (A) NAME / KEY: Coding Sequence

     (B) LOCATION: 1......6850

25   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:33:
     gacggatcggggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
     agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
     acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
     gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
30   gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
     cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
     accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
     gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
     agcccggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
35   acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
     tggtactccgtctcagaggtgtgcaccacggatcctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
     gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
     aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
     cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
40   ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgaattcgccaccatggtgagcaagggcgaggagctgt
     tcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctg
     aagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacatgaagcagcac
     gacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacacc
     ctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaaggtctatatcaccgccga
45   caagcagaagaacggcatcaaggtgaacttcaagacccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacgg
     ccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagacccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccg
     ggatcactctcggcatggacgagctgtacaagtaagtctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgcca
     gccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcatt
     ctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaaataggaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaacc
50   agctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgccc
     gctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgacccc
     aaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttcttaatagtggactcttgttccaaactg
     gaacaacactcaaccctatctcggtctattctttgatttataagggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattc
     tgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcccag
55   gctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattc

35

tccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggctttttgcaaa
aagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgccgcttgggtg
gagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtc
cggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgtctcgacgttgtcactgaagcgggaagggactg
gctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccgg
ctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcg
cgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgctt
ttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcc
tcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcc
caacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgc
tggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtt
tgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattc
cacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacct
gtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgc
ggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccg
taaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagata
ccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctca
cgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtcc
aacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactac
ggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtgg
tttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggat
tttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgct
taatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtg
ctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctcca
tccagtctattaattgttgccggggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatg
gcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggcc
gcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtg
tatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactct
caaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggc
aaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggataca
tatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:34:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt

agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggggttaggcgttttgcgctgcttcgcgatgtacgggccagatat

acgcgttgacattgattattgactagttattaatagtaatcaattacgggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg

gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact

gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt

cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc

accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg

gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc

agcccggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag

acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc

tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat

gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag

aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc

cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc

ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagac

catcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaaca

gctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgaattcgccaccatggtg

agcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccac

ctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccttcacctacggcgtgcagtgcttcgcccgctacccc

gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggt

gaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccac

aaggtctatatcaccgccgacaagcagaagaacggcatcaaggtgaacttcaagacccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcaga

acacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctg

gagttcgtgaccgccgcgggatcactctcggcatggacgagctgtacaagtaagtctagagggcccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcg

actgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttttcctaataaaatgaggaaattgcatcgc

attgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggc

ttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacact

tgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatcccctttagggttccgatttagtg

ctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatag

tggactcttgttccaaactggaacaacactcaacccctatctcggtctattcttttgatttataagggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaa

aatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaacc

aggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaact

ccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttg

gaggcctaggcttttgcaaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgca

ggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttct

tttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtca

ctgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcg

gctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggac

gaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatc

atggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcgg

cgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaa

tgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctc

cagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttt

cactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaa

attgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgccc

gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgct

gcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggcc

agcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaac

ccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggataccgtccgcctttctcccttcgggaagc

gtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatcc

ggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttct

tgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaa

accaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaac

gaaaactcacgttaagggatttttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttaaatcaatctaaagtatatatgagtaaactt
ggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggaggg
cttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcc
tgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgt

5   cacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgt
tgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacca
agtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgt
tcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttactttcaccagcgtttctgggt
gagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttat

10 tgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.35: (Caspase 3)

15   (i) SEQUENCE CHARACTERISTICS:

    (A) LENGTH: 834 base pairs

    (B) TYPE: nucleic acid

20

    (C) STRANDEDNESS: double

    (D) TOPOLOGY: linear

25   (ii) MOLECULE TYPE: cDNA

    (ix) FEATURE:

    (A) NAME / KEY: Coding Sequence

30

    (B) LOCATION: 1......834

    (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:35:
atggagaacactgaaaactcagtggattcaaaatccattaaaaatttggaaccaaagatcatacatggaagcgaatcaatggactctggaatatccctggacaacagttata

35 aaatggattatcctgagatgggtttatgtgtataataattaataataagaattttcataaaagcactggaatgacatctcggtctggtacagatgtcgatgcagcaaacctcaggg
aaacattcagaaacttgaaatatgaagtcaggaataaaaatgatcttacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagcaaaaggagcag
ttttgtttgtgtgcttctgagccatggtgaagaaggaataattttttggaacaaatggacctgttgacctgaaaaaaataacaaactttttcagaggggatcgttgtagaagtcta
actggaaaacccaaacttttcattattcaggcctgccgtggtacagaaactggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaataccagt
ggaggccgacttcttgtatgcatactccacagcacctggttattattcttggcgaaattcaaaggatggctcctggttcatccagtcgctttgtgccatgctgaaacagtatgcc

40 gacaagcttgaatttatgcacattcttacccgggttaaccgaaaggtggcaacagaatttgagtccttttcctttgacgctacttttcatgcaaagaaacagattccatgtattg
tttccatgctcacaaaagaactctattttatcactaa

2) INFORMATION FOR SEQ. ID. NO.36: (C35' primer)

45   (i) SEQUENCE CHARACTERISTICS:

    (A) LENGTH: 51 base pairs

50   (B) TYPE: nucleic acid

    (C) STRANDEDNESS: single

    (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: oligonucleotides

(ix) FEATURE:

5　(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....51

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:36:

10

CGGATCCAACACTGAAAACTCAGTGGATTCAAAATCCATTAAAAATTTGG

(2) INFORMATION FOR SEQ. ID. NO:37:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

20　(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

30　(A) NAME / KEY: Coding Sequence

(B) LOCATION:　148

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:37:

35

CGGATCCGTGATAAAAATAGAGTTCTTTTGTGAGCATGGAAACAATAC

40

2) INFORMATION FOR SEQ. ID. NO.38: (pcDNA3-1XUb-C3)

45　(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6436 base pairs

(B) TYPE: nucleic acid

50

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

55　(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6436

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:38:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgggaaacagctggaagatggacgccaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgca
ccacggatccaacactgaaaactcagtggattcaaaatccattaaaaaatttggaaccaaagatcatacatggaagcgaatcaatggactctggaatatccctggacaacagt
tataaaatggattatcctgagatgggttttatgtataataattaataataagaattttcataaaagcactggaatgacatctcggtctggtacagatgtcgatgcagcaaacctca
gggaaacattcagaaacttgaaatatgaagtcaggaataaaaatgatcttacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagcaaaaggag
cagttttgtttgtgtgcttctgagccatggtgaagaaggaataattttttggaacaaatggacctgttgacctgaaaaaaataacaaacttttttcagaggggatcgttgtagaagt
ctaactggaaaacccaaacttttcattattcaggcctgccgtggtacagaactggactgtgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaatacc
agtggaggccgacttcttgtatgcatactccacagcacctggttattattcttggcgaaattcaaaggatggctcctggttcatccagtcgctttgtgccatgctgaaacagtat
gccgacaagcttgaatttatgcacattcttacccggggttaaccgaaaggtggcaacagaatttgagtcctttcctttgacgctactttcatgcaaagaaacagattccatgta
ttgtttccatgctcacaaaagaactctatttttatcacggatcctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagtt
gccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtg
tcattctattctggggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaag
aaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagc
gcccgctccttcgctttcttcccttccttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcga
ccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaa
actggaacaacactcaaccctatctcggtctattctttgatttataagggattttggggattcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatta
attctgtggaatgtgtgtcagttaggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcc
ccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcc
cattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttg
caaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgctt
gggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccga
cctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaag
ggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttg
atccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcagg
ggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatg
gccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgac
cgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaag
cgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggat
ctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctag
ttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgct
cacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcg
ggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgt
tcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggcc
aggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggacta
taaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctc
aatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtc
ttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcc
taactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggta
gcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgtt

aagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttac
caatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcc
ccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccg
cctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgttt

5   ggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaag
ttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgaga
atagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaa
aactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacag
gaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcg

10   gatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:39:

15   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6703 base pairs

(B) TYPE: nucleic acid

20

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

25   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

30

(B) LOCATION: 1......6703

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:39:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt

35   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc

40   acccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agccccggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat

45   gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccaacactgaaaactcagtggattcaaaatccattaaaaatttggaaccaaagatcatac
atggaagcgaatcaatggactctggaatatccctggacaacagttataaaatggattatcctgagatgggtttatgtataataattaataataagaattttcataaaagcactg
gaatgacatctcggtctggtacagatgtcgatgcagcaaacctcagggaaacattcagaaacttgaaatatgaagtcaggaataaaaatgatcttacacgtgaagaaattgt
ggaattgatgcgtgatgtttctaaagaagatcacagcaaaaggagcagttttgtttgtgtgcttctgagccatggtgaagaaggaataattttttggaacaaatggacctgttga

50   cctgaaaaaaataaacaacttttcagagggggatcgttgtagaagtctaactggaaaacccaaactttcattattcaggcctgccgtggtacagaactggactgtggcattg
agacagacagtggtgttgatgatgacatggccgtgtcataaaataccagtggaggccgacttcttgtatgcatactccacagcacctggttattattcttggcgaaattcaaagg
atggctcctggttcatccagtcgctttgtgccatgctgaaacagtatgccgacaagcttgaatttatgcacattcttacccgggttaaccgaaaggtggcaacagaatttgagt
ccttttccttgacgctactttcatgcaaagaaacagattccatgtattgtttccatgctcacaaaagaactctattttttatcacggatcctagagggccctattctatagtgtca
cctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatcctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtc

55   ctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagca

```
ggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtg
tggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaa
tcggggcatccctttaggggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttttcgc
cctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttggggatttcggcc
tattggttaaaaaatgagctgatttaacaaaaattaacgcgaattaattctgtgggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtat
gcaaagcatgcatctcaattagtcagcaaccaggtgtgtgaaagtccccaggctcccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtccc
gcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctga
gctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatcgtgatcaagagacaggatgaggatcgtt
tcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgt
tccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacggg
cgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgaga
aagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatgga
agccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgt
gacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttgg
ctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacg
agttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaa
tcgttttccggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaata
gcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttgg
cgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagct
aactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggg
cgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggggataa
cgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaa
atcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccgg
atacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccc
cccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagag
cgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaa
agagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgat
cttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttt
aaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccc
cgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagcca
gccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaa
cgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaa
aaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaa
gatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagc
agaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatct
tcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcc
tttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtg
ccacctgacgtc
```

(2) INFORMATION FOR SEQ. ID. NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6946 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6946

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:40:

```
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaaccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccggggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccaacactgaaaactcagtggattcaaaatccattaaaa
atttggaaccaaagatcatacatggaagcgaatcaatggactctggaatatccctggacaacagttataaaatggattatcctgagatgggtttatgtataataattaataata
agaattttcataaaagcactggaatgacatctcggtctggtacagatgtcgatgcagcaaacctcagggaaacattcagaaacttgaaatatgaagtcaggaataaaaatga
tcttacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagcaaaaggagcagtttttgtttgtgtgctctgagccatggtgaagaaggaataattttt
ggaacaaatggacctgttgacctgaaaaaaaataacaaacttttttcagagggggatcgttgtagaagtctaactggaaaacccaaactttcattattcaggcctgccgtggtac
agaactggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaataccagtggaggccgacttcttgtatgcatactccacagcacctggttatt
attcttggcgaaattcaaaggatggctcctggttcatccagtcgctttgtgccatgctgaaacagtatgccgacaagcttgaatttatgcacattcttacccgggttaaccgaa
aggtggcaacagaatttgagtccttttcctttgacgctactttcatgcaaagaaacagattccatgtattgtttccatgctcacaaaagaactctatttttatcacggatcctag
agggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttcagttgccagccatctgttgtttgccctccccgtgccttccttgaccctg
gaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtgggtgggcaggacagcaaggggga
ggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctgggggctctagggggtatcccacgcgccctgtagcgg
cgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttcgctttcttcccttcctttctcgccacgttcgccgg
ctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatc
gccctgatagacggtttttcgcccttgacgttggagtccacgttcttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttata
agggattttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggggtgtggaaagtccccaggct
ccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaat
tagtcagcaaccatagtcccgcccctaactccgcccatccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggcc
gaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaa
gagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcg
gctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggct
atcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtca
tctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcga
gcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcc
cgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccgcctgggtgtggcggacc
gctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatc
gccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttct
atgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttata
atggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgt
cgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcct
ggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcgggg
agaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggtta
tccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccc
ccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgtt
ccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagc
tgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccac
```

tggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctga
agccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaagg
atctcaagaagatcctttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctt
ttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttc
atccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagat
ttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttc
gccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttaca
tgatccccCatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctc
ttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacggg
ataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaaccc
actcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaa
tgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccg
cgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7189 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......7189

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:41:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggggttaggcgtttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagac
catcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaaca
gctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccaacactgaaaactcagt
ggattcaaaatccattaaaaatttggaaccaaagatcatacatggaagcgaatcaatggactctggaatatccctggcaacagttataaaaatggattatcctgagatgggtt
tatgtataataattaataataagaattttcataaaagcactggaatgacatctcggtctgtacagatgtcgatgcagcaaacctcaggggaaacattcagaaacttgaaatat

gaagtcaggaataaaaatgatcttacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagcaaaaggagcagttttgtttgtgtgcttctgagccat
ggtgaagaaggaataattttttggaacaaatggacctgttgacctgaaaaaaataacaaactttttcagagggggatcgttgtagaagtctaactggaaaacccaaactttcat
tattcaggcctgccgtggtacagaactggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcataaaataccagtggaggccgacttcttgtatgcat
actccacagcacctggttattattcttggcgaaattcaaaggatggctcctggttcatccagtcgctttgtgccatgctgaaacagtatgccgacaagcttgaattatgcacat
tcttacccgggttaaccgaaaggtggcaacagaatttgagtcctttttccttgacgctactttcatgcaaagaaacagattccatgtattgtttccatgctcacaaaagaactc
tattttttatcacggatcctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccctcc
cccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtgg
ggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtat
ccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttcgctttcttcccttc
ctttctcgccacgttcgccggctttcccgtcaagctctaaatcggggcatccctttagggttccgatttagtgcttttacggcacctcgaccccaaaaaacttgattagggtgat
ggttcacgtagtgggccatcgccctgatagacggttttttcgcccttttgacgttggagttccacgttcttttaatagtggactcttgttccaaactggaacaacactcaacccatct
cggtctattcttttgatttataagggatttttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagg
gtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagt
atgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaat
ttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtata
tccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatga
ctgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgc
aggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgc
cggggcaggatcctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccac
caagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgcc
aggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgctttctggattcatcgactgtgg
ccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccg
ctccccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctgggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatt
tcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccca
acttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtat
cttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccgg
aagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaat
gaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcac
tcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctgg
cgttttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaa
gctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttc
ggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgact
tatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagt
atttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagca
gattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaa
aaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctat
ctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgaga
cccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgcc
gggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc
ccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggt
tatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgc
tcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgtt
gagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaag
ggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaa
aaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.42: C35'Met primer

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 54 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

5   (ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1....53

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:42:

15

CGGATCCATGAACACTGAAAACTCAGTGGATTCAAAATCCATTAAAAATTTGG

20
2) INFORMATION FOR SEQ. ID. NO.43: C33' primer

(i) SEQUENCE CHARACTERISTICS:

25   (A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

30   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

35   (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 148

40   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:43:

45   CGGATCCGTGATAAAAATAGAGTTCTTTTGTGAGCATGGAAACAATAC

2) INFORMATION FOR SEQ. ID. NO.44: pcDNA3-Ub-Met-C3

50   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7248 base pairs

(B) TYPE: nucleic acid

55

46

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

5   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10

(B) LOCATION: 1......7248

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:44:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
15   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggacttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
20   accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgggat
gcacggatcccacatcaacactgaaaactcagtggcctcaaaatccattaaaaaatttggaaccaaagatcatacatggaagcgaatcaatggactctggaatatccctggac
25   aacagttataaaatggattatcctgagatgggtttatgtataataattaataataagaattttcataaaagcactggaatgacatctcggtctggtacagatgtcgatgcagca
aacctcagggaaacattcagaaacttgaaatatgaagtcaggaataaaaatgatcttacacgtgaagaaattgtggaattgatgcgtgatgtttctaaagaagatcacagca
aaaggagcagttttgtttgtgtgcttctgagccatggtgaagaaggaataattttttggaacaaatggacctgttgacctgaaaaaaataacaaacttttcagaggggatcgtt
gtagaagtctaactggaaaaacccaaacttttcattattcaggcctgccgtggtacagaactggactgtggcattgagacagacagtggtgttgatgatgacatggcgtgtcat
aaaataccagtggatgccgacttcttgtatgcatactccacagcacctggttattattcttggcgaaattcaaaggatggctcctggttcatccagtcgctttgtgccatgctga
30   aacagtatgccgacaagcttgaatttatgcacattcttacccgggttaaccgaaaggtggcaacagaatttgagtcctttccttttgacgctactttcatgcaaagaaacaga
ttccatgtattgtttccatgctcacaaaagaactctattttttatcacggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtg
cacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcgg
tattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatg
acagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaaca
35   tggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgca
aactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctgg
tttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtca
ggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgc
tgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgagg
40   aaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtgggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggt
gggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgt
gaccgctacacttgccagcgccctagcgcccgctccttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttaggg
ttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtcca
cgttctttaatagtggactcttgttccaaactggaacaacactcaacccctatctcggtctattcttttgatttataagggattttgggattcggcctattggttaaaaaatgagc
45   tgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaat
tagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatc
ccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgag
gaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatg
gattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggg
50   gcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtg
ctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctga
tgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcag
gatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctg
cttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctga
55   agagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggact

ctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccg
gctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaa
ataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagc
tgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgt
5   tgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcg
ctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgt
gagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcag
aggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttct
cccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccg
10 ctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggt
gctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttga
tccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgac
gctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttaaatcaatctaaagtatat
atgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactac
15 gatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagc
gcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctac
aggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcctt
cggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactg
gtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctc
20 atcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttca
ccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaag
catttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtc

25
2) INFORMATION FOR SEQ. ID. NO.45: DEVD-1 primer

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
35
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

40 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....48
45
(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO: 45:

GATCCGTCGGCGCTGTCGGCAGCGTCGGCGACGAGGTCGACGGCGTCG

50
(2) INFORMATION FOR SEQ. ID. NO.: 46:

(i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

5

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

10 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1....48

15

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:46:
GATCCGACGCCGTCGACCTCGTCGCCGACGCTGCCGACAGCGCCGACG

20

2) INFORMATION FOR SEQ. ID. NO.47: pcDNA3-1XUb-DEVD-Bla

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 6459 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

35 (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6459

40

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:47:
gacggatcggagagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
45 gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
acccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
50 tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgca
ccacggatccgtcggcgctgtcggcagcgtcggcgacgaggtcgacggcgtcggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatc
agttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtttttcgccccgaagaacgtttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacg
55 gatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttt

49

tgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaa
cgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccg
gctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggcagatggtaagccctcccgtatcgtagttatctacacgacg
gggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgcta
5 gagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaata
aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcagcaagggggaggattgggaagacaatagcaggcatgctgg
ggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttac
gcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttttcgctttcttcccttccttttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatc
cctttaggggttccgatttagttgcttttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgcccctgatagacggttttttcgccctttgacgt
10 tggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaacccctatctcggtctattcttttgatttataaggggattttgggggatttcggcctattggttaa
aaaatgagctgatttaacaaaaatttaacgcgaattaattctgtgtgaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcat
gcatctcaattagtcagcaaccaggtgtggaaagtcccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaact
ccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccggaggccgcctctgcctctgagctattccaga
agtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttttcggatctgatcaagagacaggatgaggatcgtttcgcatgattg
15 aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtc
agcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgc
gcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccat
catggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtctt
gtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatgg
20 cgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtg
atattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctg
agcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccg
ggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccacccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa
atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcat
25 ggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcaca
ttaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttcc
gcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaa
agaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgc
tcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtc
30 cgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcag
cccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatg
taggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggt
agctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatctttctacg
gggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct
35 aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtag
ataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagg
gccgagcgcagaagtggtcctgcaacttttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgcc
attgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggtt
agctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttc
40 tgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaa
aagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctt
ttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatat
tattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgac
gtc
45

(2) INFORMATION FOR SEQ. ID. NO:48:

(i) SEQUENCE CHARACTERISTICS:
50
(A) LENGTH: 6726 base pairs

(B) TYPE: nucleic acid
55 (C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6726

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:48:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttcgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgctgtcggcagcgtcggcgacgaggtcgacggcgtcggatccggggcgtgg
ctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgcc
ccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcaga
atgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaactt
acttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacca
aacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatgg
aggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactg
gggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagc
attggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgc
cttccttgaccctggaaggtgccactcccactgtccttttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcagg
acagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccac
gcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctc
gccacgttcgccggctttccccgtcaagctctaaatcgggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttca
cgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtct
attctttgatttataagggattttgggggtttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtgg
aaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaa
agcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttt
atttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccatt
ttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggc
acaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggac
gaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccgggg
caggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagc
gaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggct
caaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggc
tgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctccc
gattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgat
tccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttg
tttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttat
catgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagc
ataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatc
ggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaa

ggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttt
tccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccc
tcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtag
gtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcc
5   actggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggt
atctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacg
cgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggat
cttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcg
atctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgc
10   tcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaag
ctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacg
atcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggc
agcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgc
ccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatc
15   cagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaata
agggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataa
acaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

gacggatcgggagatctcccgatccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
45   acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
50   gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
55   aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc

cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc

ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgctgtcggcagcgtcggcgacgaggtcgac

ggcgtcggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggt

aagatccttgagagttttcgccccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggt

cgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgag

tgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgg

agctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggc

aacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggttattgctgataaatctggagccggtgagcgtgggtct

cgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgag

ataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatct

gttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattct

ggggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctgg

ggctctaggggggtatcccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctt

tcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaa

cttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaac

actcaaccctatctcggtctattctttgatttataaggggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaat

gtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctcccc

agcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccc

catggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcc

cgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagagg

ctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgc

cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgct

attgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacct

gcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgcca

gccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctg

gattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtg

ctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcggggactctggggttcgaaatgaccgaccaagcgacgcccaac

ctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctgga

gttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttttcactgcattctagttgtggtttgtc

caaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccac

acaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtc

gtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggc

gagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaa

aaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca

ggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgc

tgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaac

ccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggct

acactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttt

ttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttg

gtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaat

cagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgc

aatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatcca

gtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggctt

cattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgca

gtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtat

gcgcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctca

aggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaa

aatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatat

ttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7212 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME /.KEY: Coding Sequence

(B) LOCATION: 1......7212

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:50:

```
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttlaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgaccccgccccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccggggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagagccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagac
catcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaaca
gctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgcgtcggcag
cgtcggcgacgaggtcgacggcgtcggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcga
actggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgcc
gggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgca
gtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatggggggatcatgtaactcg
ccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaact
acttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggcccattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgcc
ttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctga
gtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtggcctctatggcttctgaggc
ggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgc
cctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggca
cctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttg
ttccaaactggaacaacactcaaccctatctcggtctattctttgatttataagggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc
gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtgga
aagtcccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagtt
ccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctagg
cttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggc
cgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaaga
ccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggg
```

aagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacg
cttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcat
caggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaa
aatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggc
tgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctgggggttcgaaatgaccgacc
aagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttggcttcggaatcgtttttccgggacgccggctggatgatcctccagcgcgg
ggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttttcactgcatt
ctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatc
cgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccag
tcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggt
cgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaagg
ccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacagga
ctataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgcttt
ctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatc
gtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtg
gcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctg
gtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcac
gttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagt
taccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctg
gccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttat
ccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtc
gtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagt
aagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg
agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggc
gaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctttactttcaccagcgtttctgggtgagcaaaaa
caggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatga
gcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......48

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:51:
GATCCGTCGGCGCTGTCGGCAGCGTCGGCGACGAGGTCGCTGGCGTCG

(2) INFORMATION FOR SEQ. ID. NO:52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......48

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:52:
GATCCGACGCCAGCGACCTCGTCGCCGACGCTGCCGACAGCGCCGACG

(2) INFORMATION FOR SEQ. ID. NO:53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6459 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6459

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:53:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgaccccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc

agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgca
ccacggatccgtcggcgctgtcggcagcgtcggcgacgaggtcgctggcgtcggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatc
agttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctat
gtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacg
datggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttt
tgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaa
cgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccg
gctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacg
gggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgcta
gagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaata
aaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctgg
ggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttac
gcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttcgctttcttcccttccttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggcatc
cctttaggggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgt
tggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaacccctatctcggtctattctttgatttataaggggattttgggggatttcggcctattggttaa
aaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcat
gcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaact
ccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccaga
agtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattg
aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtc
agcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgc
gcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatcctcctgtcatctcaccttgctcctgccgagaaagtatccat
catggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtctt
gtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatgg
cgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtg
atattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctg
agcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccg
ggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccacccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa
atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcat
ggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcaca
ttaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttcc
gcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaa
agaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgc
tcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtc
cgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcag
cccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatg
taggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggt
agctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacg
gggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct
aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtag
ataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagg
gccgagcgcagaagtggtcctgcaacttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgcc
attgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggtt
agctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttc
tgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaa
aagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatctt
ttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttaatactcatactcttcctttttcaatat
tattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgac
gtc

(2) INFORMATION FOR SEQ. ID. NO:54:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6726 base pairs

5   (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

10

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

15   (A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6726

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:54:

20   gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacgggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt

25   cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaaaccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccggggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc

30   tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgctgtcggcagcgtcggcgacgaggtcgctggcgtcggatccggggcgtgg
ctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtttcgcc
ccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcaga

35   atgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaactt
acttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacca
aacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatgg
aggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactg
gggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagc

40   attggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgc
cttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcagg
acagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccac
gcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctc
gccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttca

45   cgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtct
attctttttgatttataagggattttgggattttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtgg
aaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaa
agcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttt
atttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccatt

50   ttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggc
acaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggac
gaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccgggg
caggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagc
gaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggct

55   caaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggc

58

tgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctccc
gattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgat
tccaccgccgccttctatgaaaggttgggcttcggaatcgtttccgggacgccggctggatgatcctccagcgcgggggatctcatgctggagttcttcgcccaccccaacttg
tttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttat

catgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagc
ataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatc
ggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaa
ggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttt
tccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccc

tcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtag
gtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcc
actggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggt
atctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacg
cgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggat

cttcacctagatcctttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcg
atctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgc
tcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaag
ctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacg
atcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggc

agcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgc
ccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatc
cagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaata
agggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataa
acaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6969 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......6969

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:55:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgtttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactt
tcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc

agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag

aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgctgtcggcagcgtcggcgacgaggtcgctg
gcgtcggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta
agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtc

gccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagt
gataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgg
agctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggc
aacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggccccttccggctggctggttatttgctgataaatctggagccggtgagcgtgggtct
cgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgag

ataggtgcctcactgattaagcattggtaatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatct
gttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattct
gggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctgg
ggctctaggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctt
tcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgacccccaaaaaa

cttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttttttcgccctttgacgttggagtccacgttcttttaatagtggactcttgttccaaactggaacaac
actcaacccctatctcggtctattctttgatttataaggggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaat
gtgtgtcagttagggtgtggaaagtcccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcccaggctcccc
agcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccc
catggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcc

cgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagagg
ctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgc
cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgct
attgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacct
gcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcagggggctcgcgcca

gccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctg
gattcatcgactgtggccggctgtgggtgtggcggaccgctatcaggacatagcgttggctaccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtg
ctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaac
ctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctgga
gttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtc

caaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagacttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccac
acaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtc
gtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggc
gagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaa
aaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca

ggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgc
tgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaac
ccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggct
acactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttt
ttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttg

gtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaat
cagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgc
aatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatcca
gtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggctt
cattcagctccggttcccaacgatcaaggcgagttacatgatccccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgca

gtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtat
gcggcgaccgagttgctcttgcccggcgtcaatacggggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctca
aggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaa
aatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatat
ttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7212 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......7212

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:56:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgaccccgccattgacgtcaataatgacgtatgttcccatagtaacgccaataagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagac
catcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaaca
gctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccgtcggcgctgtcggcag
cgtcggcgacgaggtcgctggcgtcggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcga
actggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgcc
gggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgca
gtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcg
ccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaact
acttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccctattctatatgtgtcacctaaatgctagagctcgctgatcagcctcgactgtgcc
ttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctga
gtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggc
ggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgc
cctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatcccttagggttccgatttagtgctttacggca
cctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttg
ttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgggaatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgc

gaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtgga
aagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagtt
ccgcccattctccgccccatggctgactaattttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctagg
cttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggc
cgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaaga
ccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggg
aagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacg
cttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcat
caggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaa
aatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggc
tgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgacc
aagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgtttccgggacgccggctggatgatcctccagcgcgg
ggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttttcactgcatt
ctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatc
cgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccag
tcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggt
cgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaagg
ccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacagga
ctataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgcttt
ctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatc
gtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtg
gcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctg
gtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcac
gttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagt
taccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctg
gccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttat
ccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtc
gtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagt
aagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctg
agaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggc
gaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaa
caggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcaggggttattgtctcatga
gcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.57: rhinovirus 14 2a

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1095 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......1095

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:57:
ttgggtcgtgcagcttgtgtgcatgtaactgaaatacaaaacaaagatgctactggaatagataatcacagagaagcaaaattgttcaatgattggaaaatcaacctgtccag
ccttgtccaacttagaaagaaactggaactcttcacttatgttaggtttgattctgagtataccatactggccactgcatctcaacctgattcagcaaactattcaagcaatttg
gtggtccaagccatgtatgttccacatggtgccccgaaatccaaaagagtgggcgattacacatggcaaagtgcttcaaaccccagtgtattcttcaaggtgggggatacatc
aaggtttagtgtgccttatgtaggattggcatcagcatataattgtttttatgatggttactcacatgatgatgcagaaactcagtatggcataactgttctaaaccatatgggta
gtatggcattcagaatagtaaatgaacatgatgaacacaaaactcttgtcaagatcagagtttatcacagggcaaagctcgttgaagcatggattccaagagcacccagagc
actaccctacacatcaatagggcgcacaaattatcctaagaatacagaaccagtaattaagaagaggaaaggtgacattaaatcctatggtttaggacctaggtacggtggg
atttatacatcaaatgttaaaataatgaattaccacttgatgacaccagaagaccaccataatctgatagcaccctatccaaatagagatttagcaatagtctcaacaggagg
acatggtgcagaaacaataccacactgtaaccgtacatcaggtgtttactattccacatattacagaaagtattaccccataatttgcgaaaagcccaccaacatctggattg
aaggaagcccttattacccaagtagatttcaagcaggagtgatgaaaggggttgggccggcagagctaggagactgcggtgggattttgagatgcatacatggtcccattgg
attgttaacagctgaaggtagtggatatgtttgttttgctgacatacgacagttggagtgtatcgcagaggaacag

2) INFORMATION FOR SEQ. ID. NO.58: HRV145' primer

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:58:
taggatccttgggtcgtgcagcttgtgtg

(2) INFORMATION FOR SEQ. ID. NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 nubleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:59:
aaggatccctgttcctctgccatacactc

(2) INFORMATION FOR SEQ. ID. NO:60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8022 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......8022

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:60:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccttgggtcgtgcagcttgtgtgcatgtaactgaaataca
aaacaaagatgctactggaatagataatcacagagaagcaaaattgttcaatgattggaaaatcaacctgtccagccttgtccaacttagaaagaaactggaactcttcactt
atgttaggtttgattctgagtataccatactggccactgcatctcaacctgattcagcaaactattcaagcaatttggtggtccaagccatgtatgttccacatggtgccccgaa
atccaaaagagtgggcgattacacatggcaaagtgcttcaaaccccagtgtattcttcaaggtgggggatacatcaaggtttagtgtgccttatgtaggattggcatcagcat
ataattgtttttatgatggttactcacatgatgatgcagaaactcagtatggcataactgttctaaaccatatgggtagtatggcattcagaatagtaaatgaacatgatgaaca
caaaactcttgtcaagatcagagtttatcacagggcaaagctcgttgaagcatggattccaagagcacccagagcactaccctacacatcaatagggcgcacaaattatcct
aagaatacgaaccagtaattaagaagaggaaaggtgacattaaatcctatggtttaggacctaggtacggtgggatttatacatcaaatgttaaaataatgaattaccactt
gatgacaccagaagaccaccataatctgatagcaccctatccaaatagagatttagcaatagtctcaacaggaggacatggtgcagaaacaataccacactgtaaccgtac
atcaggtgtttactattccacatattacagaaagtattaccccataatttgcgaaaagcccaccaacatctggattgaaggaagccettattacccaagtagatttcaagcagg
agtgatgaaagggggttgggccggcagagctaggagactgcggtgggattttgagatgcatacatggtcccattggattgttaacagctgaaggtagtggatatgtttgttttgc
tgacatacgacagttggagtgtatcgcagaggaacagggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagt
gggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgcccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcc

cgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaa

gagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatggggga

tcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactatta

actggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgct

gataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactgggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaacta

tggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggcccctattctatagtgtcacctaaatgctagagctcgctgatcagc

ctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcat

cgcattgtctgagtaggtgtcattctattctggggggtgggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctat

ggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac

acttgccagcgccctagcgcccgctccttttcgctttcttccccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgattta

gtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccctttgacgttggagtccacgttctttaa

tagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggatttgggattcggcctattggttaaaaaatgagctgatttaac

aaaaattaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtcccaggcctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagca

accaggtgtggaaagtcccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgccccta

actccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttt

ttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacg

caggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggt

tcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttg

tcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatcctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcg

gcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctg

gacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaat

atcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggc

ggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcga

aatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctgggatgatc

ctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttt

tttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtg

aaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgc

ccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcg

ctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggc

cagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaa

cccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaag

cgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatc

cggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttc

ttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaca

aaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaa

cgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaact

tggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagg

gcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtc

ctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggt

gtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatc

gttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaac

caagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaac

gttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgg

gtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggtt

attgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.61: rhinovirus 16 2a

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 636 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10

(B) LOCATION: 1......636

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:61:
atgggaactttgtgttcgcgtattgtgaccagtgagcaattacacaaagtcaaagtggtaacaaggatatatcacaaagccaaacacaccaaagcttggtgcccaagaccac

15   ccagagctgttcaatactcacatacacataccaccaactacaaattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactgttgggcctagt
gacatgtatgtgcatgttggtaatctaatatacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaatcatataccgaacaagc
acacaaggtgatggttatattccaacatgtaattgcactgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgactggtatgagat
acaagagagtgaatattatccaaaacatatccagtacaatttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacatggagtgatag
gtattattacagcaggtggtgagggccatgttgcattcatagatcttagacactttcactgtgctgaa

20

2) INFORMATION FOR SEQ. ID. NO.62: HRV165 primer

(i) SEQUENCE CHARACTERISTICS:
25

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

30   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide
35

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

40   (B) LOCATION: 1......29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:62:

aaggatccatgggaactttgtgttcgcgt

45

(2) INFORMATION FOR SEQ. ID. NO:63:

(i) SEQUENCE CHARACTERISTICS:
50

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

55   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

5

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

10   (B) LOCATION: 1......29

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:63:
ttggatccttcttcagcacagtgaaagtgtc

15

(2) INFORMATION FOR SEQ. ID. NO:64:

20   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7563 base pairs

(B) TYPE: nucleic acid

25

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

30   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

35

(B) LOCATION: 1......7563

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:64:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
40   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
45   accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
50   gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacccgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccatgggaactttgtgttcgcgtattgtgaccagtgagca
attacacaaagtcaaagtggtaacaaggatatatcacaaagcccaaacacaccaaagcttggtgcccaagaccacccagagctgttcaatactcacatacacataccaccaa
55   ctacaaattgagttcagaagtacacaatgatgtgtggctataagacctagaacaaatctaacaactgttgggcctagtgacatgtatgtgcatgttggtaatctaatatacagaaa

67

tctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaatcatataccgaacaagcacacaaggtgatggttatattccaacatgtaattgcac

tgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgactggtatgagatacaagagagtgaatattatccaaaacatatccagtac

aatttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacatggagtgataggtattattacagcaggtggtgagggccatgttgcattc

atagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttaca

tcgaactggatctcaacagcggtaagatccttgagagtttcgccccgaagaacgtttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattga

cgccgggcaagagcaactcggtcgccgccatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaatta

tgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaa

ctcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcg

aactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaat

ctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatga

acgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggcccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgact

gtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattg

tctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttct

gaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc

agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgcttta

cggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtgga

ctcttgttccaaactggaacaacactcaacccctatctcggtctattcttttgatttataaggggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaattt

aacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggt

gtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc

ccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggagg

cctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatccgtttcgcatgattgaacaagatggattgcacgcaggttc

tccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttg

tcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactga

agcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcacctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctg

catacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaa

gagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatg

gtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcga

atgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatga

ccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccag

cgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcac

tgcattcagttggtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaatt

gttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgct

ttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcg

ctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagc

aaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccg

acaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtg

gcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggt

aactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttga

agtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacc

accgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaa

aactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtc

tgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggctta

ccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgc

aactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtca

cgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttg

tcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaa

gtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgtt

cttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtg

agcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcaggggtatt

gtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

(2) INFORMATION FOR SEQ. ID. NO:65:

(i) SEQUENCE CHARACTERISTICS:

5   (A) LENGTH: 7053 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

10

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: cDNA

15   (ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......7053

20

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:65:

```
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaataagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgggat
gcacggatccatgggaactttgtgttcgcgtattgtgaccagtgagcaattacacaaagtcaaagtggtaacaaggatatatcacaaagccaaacacaccaaagcttggtgc
ccaagaccacccagagctgttcaatactcacatacacataccaccaactacaaattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactg
ttgggcctagtgacatgtatgtgcatgttggtaatctaatatacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaatcatata
ccgaacaagcacacaaggtgatggttatattccaacatgtaattgcactgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgact
ggtatgagatacaagagagtgaatattatccaaaacatatccagtacaatttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacat
ggagtgataggtattattacagcaggtggtgagggccatgttgcattcatagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacccagaaacgctggt
gaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtttctgccccgaagaacgttttccaatga
tgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcac
cagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccac
gatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcagga
ccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcc
cgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccct
attctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccctcccccgtgccttccttgaccctggaaggtgc
cactcccactgtccttctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtggggtggggcaggacagcaaggggggaggattggg
aagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggtatccccacgcgccctgtagcggcgcattaa
gcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttcgctttcttcccttccttctcgccacgttcgccggctttccccg
tcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgata
gacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattctttgatttataagggattt
gggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggggtgtggaaagtccccaggctccccaggc
aggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagc
aaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgc
ctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacagg
atgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctg
```

atgccgccgtgttccggctgtcagcgcagggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggct
ggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcacctt
gctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagca
cgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggc
gaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatca
ggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttcta
tcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaag
gttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttac
aaataaagcaatagcatcacaaatttcacaaataaagcattttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctct
agctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgc
ctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggc
ggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccaca
gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctga
cgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgac
cctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggc
tgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggta
acaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagcca
gttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctca
agaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttaaat
taaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccat
agttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca
gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccag
ttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatc
ccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttact
gtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataat
accgcgccacatagcagaacttttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcg
tgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttga
atactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcac
atttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.66: HRV16 D35A primer

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

(B) LOCATION: 1......32

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:66:
gtgtcttattcatcagctttaatcatataccg

(2) INFORMATION FOR SEQ. ID. NO:67:

5   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 base pairs

(B) TYPE: nucleic acid

10   (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15   (ii) MOLECULE TYPE: oligonucleotide

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence

20   (B) LOCATION: 1......34

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:67:
gtgaaccaggtgatgctggtgggaaattattatg

25

2) INFORMATION FOR SEQ. ID. NO.68: pcDNA3-3XUb-Bla HRV16 (C106A)

30   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7563 base pairs

(B) TYPE:

35   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

40   (xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.68:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
45   gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacggggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
50   acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
55   ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccatgggaactttgtgttcgcgtattgtgaccagtgagca

71

attacacaaagtcaaagtggtaacaaggatatatcacaaagccaaacacaccaaagcttggtgcccaagaccacccagagctgttcaatactcacatacacataccaccaa
ctacaaattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactgttgggcctagtgacatgtatgtgcatgttggtaatctaatatacagaaa
tctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaatcatataccgaacaagcacacaaggtgatggttatattccaacatgtaattgcac
tgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgactggtatgagatacaagagagtgaatattatccaaaacatatccagtac
aatttactaataggtgaaggaccatgtgaaccaggtgatgctggtgggaaattattatgcaaacatggagtgatataggtattattacagcaggtggtgagggccatgttgcattc
atagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttaca
tcgaactggatctcaacagcggtaagatccttgagagtttcgccccgaagaacgtttccaatgatgagcactttaaagttctgctatgtggcgcggtattatcccgtattga
cgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaatta
tgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatggggggatcatgtaa
ctcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcg
aactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggttattgctgataaat
ctggagccggtgagcgtgggtctcgcggtatcattgcagcactgggggccagatggtaagccctcccgtatcgtagttatctacacgacgggggagtcaggcaactatggatga
acgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggcccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgact
gtgccttctagttgccagccatctgttgtttgccctcccccgtgccttccttgacccgtggaaggtgccactccactgtccttcctaataaaatgaggaaattgcatcgcattg
tctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttct
gaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc
agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgcttta
cggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggttttttcgccctttgacgttggagtccacgttctttaatagtgga
ctcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataaggggattttgggggattcggcctattggttaaaaaatgagctgatttaacaaaaattt
aacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggt
gtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc
ccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggagg
cctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttc
tccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttg
tcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactga
agcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctg
catacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaa
gagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatg
gtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcga
atgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatga
ccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccag
cgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcac
tgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaatt
gttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgct
ttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcg
ctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagc
aaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccg
acaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtg
gcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggt
aactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttga
agtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacc
accgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaa
aactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtc
tgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggctta
ccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgc
aactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtca
cgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttg
tcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaa
gtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgtt
cttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtg
agcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcaggtt att
gtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.69: pcDNA3-3XUb-Bla HRV16 (D35A)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7563 base pairs

(B) TYPE:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.69:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggttgactcacggggatttccaagtctcc
accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttgatatcgaattcctgc
agcccgggggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaag
acaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacc
tggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaat
gtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccag
aaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatctaccatggaaatcttcgtgaagactctgactggtaagaccatcactctcgaagtggagc
cgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgaccagcagaggttgatctttgctgggaaacagctggaagatggacgcacc
ctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgtgcaccacggatccatgggaactttgtgttcgcgtattgtgaccagtgagca
attacacaaagtcaaagtggtaacaaggatatatcacaaagccaaacacaccaaagcttggtgcccaagaccacccagagctgttcaatactcacatacacataccaccaa
ctacaaattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactgttgggcctagtgacatgtatgtgcatgttggtaatctaatatacagaaa
tctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagctttaatcatataccgaacaagcacacaaggtgatggttatattccaacatgtaattgcac
tgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgactggtatgagatacaagagagtgaatattatccaaaacatatccagtac
aatttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacatggagtgataggtattattacagcaggtggtgagggccatgttcattc
atagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttaca
tcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttccaatgatgagcactttaaagttctgctatgtggcgcggtattatcccgtattga
cgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaatta
tgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaa
ctcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcg
aactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaat
ctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatga
acgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgact
gtgccttctagttgccagccatctgttgtttgccctcccccgtgccttccttgaccctggaaggtgccactcccactgtccttcctaataaaatgaggaaattgcatcgcattg
tctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttct
gaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc
agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgcttta
cggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtgga
ctcttgttccaaactggaacaacactcaacccctatctcggtctattctttgatttataagggattttggggatttcggcctattggttaaaaaatgagctgatttaacaaaaattt
aacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggt
gtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc
ccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggagg
cctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttc
tccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttg
tcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactga
agcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctg
catacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaa

gagcatcagggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatg
gtggaaaatggccgctttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctaccgtgatattgctgaagagcttggcggcga
atgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctgggggttcgaaatga
ccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccag
cgcggggatctcatgctggagttcttcgcccaccccaacttgttttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcattttttttcac
tgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaatt
gttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgct
ttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcg
ctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagc
aaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccg
acaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtg
gcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggt
aactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttga
agtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacc
accgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaa
aactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtc
tgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggctta
ccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgc
aactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtca
cgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttg
tcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaa
gtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgtt
cttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtg
agcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttatt
gtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.70: pcDNA3-Ub-Met-Bla HRV16 (C106A)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7053 base pairs

(B) TYPE:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.70:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
acccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgtgggaaacagctggaagatggacgcaccctgtctgctacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgggat
gcacggatccatgggaactttgtgttcgcgtattgtgaccagtgagcaattacacaaagtcaaagtggtaacaaggatatatcacaaagccaaacacaccaaagcttggtgc
ccaagaccaccagagctgttcaaactcacatacacataccaccaactacaaattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactg
ttgggcctagtgacatgtatgtgcatgttggtaatctaatatacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagatttaatcatata
ccgaacaagcacacaaggtgatggttatattccaacatgtaattgcactgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgact

ggtatgagatacaagagagtgaatattatccaaaacatatccagtacaatttactaataggtgaaggaccatgtgaaccaggtgatgctggtgggaaattattatgcaaacat
ggagtgataggtattattacagcaggtggtgagggccatgttgcattcatagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacccagaaacgctggt
gaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatga
tgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcac

cagtcacagaaaagcatcttacggatggcatgacagtaagagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgctttttttgcacaacatggggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccac
gatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcagga
ccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctgcgcggtatcattgcagcactgggccagatggtaagccctcc
cgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccct

attctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgccctccccgtgccttccttgaccctggaaggtgc
cactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggggaggattggg
aagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatcccacgcgccctgtagcggcgcattaa
gcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttttcgctttcttccctcctttctcgccacgttcgccggctttccccg
tcaagctctaaatcggggcatcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgata

gacggttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggatttt
ggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggtgtggaaagtcccaggctccccaggc
aggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagc
aaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgc
ctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacagg

atgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctg
atgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggct
ggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcggggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcacctt
gctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagca
cgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggc

gaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatca
ggacatagcgttggctaccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttcta
tcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgccttctatgaaag
gttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttac
aaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctct

agctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgc
ctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggc
ggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccaca
gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctga
cgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgac

cctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggc
tgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggta
acaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagcca
gttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctca
agaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttaaat

taaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccat
agttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca
gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccag
ttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatc
ccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttact

gtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataat
accgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcg
tgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttga
atactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcac
atttccccgaaaagtgccacctgacgtc

2) INFORMATION FOR SEQ. ID. NO.71: pcDNA3-Ub-Met-Bla HRV16 (D35A)

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7053 base pairs

5   (B) TYPE:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

10

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO.71:

gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
15   gctgaccgcccaacgaccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
acccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactataggggagacccaagcttggtaccaccatggaga
20   tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgggat
gcacggatccatgggaactttgtgttcgcgtattgtgaccagtgagcaattacacaaagtcaaagtggtaacaaggatatatcacaaagccaaacacaccaaagcttggtgc
ccaagaccacccagagctgttcaatactcacatacacataccaccaactacaaattgagttcagaagtacacaatgatgtggctataagacctagaacaaatctaacaactg
ttgggcctagtgacatgtatgtgcatgttggtaatctaatatacagaaatctacatttatttaactctgacatacatgattccattttagtgtcttattcatcagctttaatcatata
25   ccgaacaagcacacaaggtgatggttatattccaacatgtaattgcactgaagctacatattactgcaaacacaaaaacaggtactacccaattaatgtcacacctcatgact
ggtatgagatacaagagagtgaatattatccaaaacatatccagtacaatttactaataggtgaaggaccatgtgaaccaggtgattgtggtgggaaattattatgcaaacat
ggagtgataggtattattacagcaggtggtgaggggccatgttgcattcatagatcttagacactttcactgtgctgaaggatccggggcgtggctgcacccagaaacgctggt
gaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatga
tgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcac
30   cagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccac
gatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcagga
ccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcc
cgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaatctagagggccct
35   attctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgc
cactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggtgtggggtggggcaggacagcaaggggggaggattggg
aagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaa
gcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttcgctttcttcccttccttctcgccacgttcgccggctttccccg
tcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgata
40   gacggttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaacccatctcggtctattctttgatttataagggattttt
ggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggtgtggaaagtccccaggctccccaggc
aggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagc
aaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgc
ctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacagg
45   atgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctg
atgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggct
ggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcacctt
gctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagca
cgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggc
50   gaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatca
ggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttcta
tcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttgattccaccgccgccttctatgaaag
gttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttac
aaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctct
55   agctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgc

ctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggc
ggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccaca
gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctga
cgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgac
5   cctgccgcttaccggatacctgtccgcctttctccctcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggc
tgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggta
acaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagcca
gttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctca
agaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcctttaaat
10   taaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccat
agttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca
gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccag
ttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatc
ccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttact
15   gtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataat
accgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcg
tgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttga
atactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcac
atttccccgaaaagtgccacctgacgtc
20

2) INFORMATION FOR SEQ. ID. NO.72: pcDNA3-MetUb-Bla HR14

25   (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7512 base pairs

(B) TYPE: nucleic acid
30
(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

35   (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME / KEY: Coding Sequence
40
(B) LOCATION: 1......7512

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:72:
gacggatcgggagatctcccgatcccctatggtcgactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagt
45   agtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatat
acgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctg
gctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaact
gcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggacttt
cctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggtttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctcc
50   accccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgg
gaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagcttggtaccaccatggaga
tcttcgtgaagactctgactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggcaaagatccaagacaaggaaggcatccctcctgacc
agcagaggttgatctttgctgggaaacagctggaagatggacgcaccctgtctgactacaacatccagaaagagtccaccctgcacctggtactccgtctcagaggtgggat
gcacggatcctgggtcgtgcagcttgtgtgcatgtaactgaaatacaaaacaaagatgctactggaatagataatcacagagaagcaaaattgttcaatgattggaaaatc
55   aacctgtccagccttgtccaacttagaaagaaactggaactcttcacttatgttaggtttgattctgagtataccatactggccactgcatctcaacctgattcagcaaactatt

caagcaatttggtggtccaagccatgtatgttccacatggtgccccgaaatccaaaagagtgggcgattacacatggcaaagtgcttcaaaccccagtgtattcttcaaggtg
ggggatacatcaaggttttagtgtgccttatgtaggattggcatcagcatataattgtttttatgatggttactcacatgatgatgcagaaactcagtatggcataactgttctaaa
ccatatgggtagtatggcattcagaatagtaaatgaacatgatgaacacaaaactcttgtcaagatcagagtttatcacagggcaaagctcgttgaagcatggattccaaga
gcacccagagcactaccctacacatcaatagggcgcacaaattatcctaagaatacagaaccagtaattaagaagaggaaaggtgacattaaatcctatggtttaggacct
aggtacggtgggatttatacatcaaatgttaaaataatgaattaccacttgatgacaccagaagaccaccataatctgatagcaccctatccaaatagagatttagcaatagt
ctcaacaggaggacatggtgcagaaacaataccacactgtaaccgtacatcaggtgtttactattccacatattacagaaagtattaccccataatttgcgaaaagccacc
aacatctggattgaaggaagccccttattacccaagtagatttcaagcaggagtgatgaaaggggttgggccggcagagctaggagactgcggtgggattttgagatgcatac
atggtcccattggattgttaacagctgaaggtagtggatatgtttgtttttgctgacatacgacagttggagtgtatcgcagaggaacagggatccggggcgttggctgcaccca
gaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaac
gtttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggt
tgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgaca
acgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgag
cgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggata
aagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagat
ggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaat
ctagagggcccctattctatagtgtcacctaaatgctagagactcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgac
cctggaaggtgccactcccactgtccttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtgggtggggcaggacagcaagg
gggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctaggggggtatccccacgcgccctgta
gcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttccttttctcgccacgttcg
ccggctttccccgtcaagctctaaatcgggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggc
catcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattctttgatt
tataagggattttgggggatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttaggggtgtggaaagtcccca
ggctccccaggcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatct
caattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagag
gccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatc
aagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaat
cggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcgg
ctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgt
catctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatc
gagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatg
cccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcgga
ccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgca
tcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccgccgcctt
ctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccacccccaacttgtttattgcagcttat
aatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccc
gtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagc
ctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcgg
ggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacgg
ttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccg
cccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcc
tgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcca
agctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagc
cactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgc
tgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaa
ggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatc
cttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcg
ttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccag
atttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagt
tcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagtta
catgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattc
tcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgg
gataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacc
cactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaa
atgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttcc
gcgcacatttccccgaaaagtgccacctgacgtc